# How the brain listens:

# Neural correlates of suprasegmental auditory processing

Thesis
presented to the Faculty of Arts
of the University of Zurich

for the degree of Doctor of Philosophy

by
Eveline Geiser

Accepted in the autumn semester 2007 on the recommendation
of
Prof. Dr. Lutz Jäncke,
and PD Dr. Valentin Marcar

Zurich 2007

*„Et quidem rhythmus sine metro esse potest, metrum vero sine rhythmo esse non potest...*
*Metrum est ratio cum modulatione, rhythmus modulatione sine ratione.“*

- Bene Venerabilis, around 700 AD -

## Acknowledgments

I first of all thank my supervisor, Dr. Martin Meyer, for giving me the opportunity to work on a topic of my choice, for his scientific guidance and mentoring and last but not least for his motivation in difficult situations. During the past three years I have truly enjoyed working in his group.

Then I would like to thank Prof. Lutz Jancke for giving me the chance to work in his laboratory. When as an undergraduate student I called him on the phone, he promptly invited me to come to his office half an hour later providing me with the possibility to talk to him. And during my postgraduate studies, he took the time to listen to my concerns whenever necessary.

Then I would like to thank the auditory group for a friendly and cheering atmosphere and their helpful comments on my work. A great thank you goes to Elsa Saegesser for linguistic comments on previous versions of this thesis.

A warm thank you goes to my parents who supported my work in so many ways and for celebrating whenever I visit them. My warmest thank you goes to Philipp for his support.

# Contents

## Summary

The present empirical work aims to elucidate neural processes underlying the perception of melody and rhythm. These suprasegmental auditory signals develop over several auditory segments and require thus the integration of information over a longer period of time. This exceeds the function of the primary auditory cortex presumably extending over secondary and higher sensory areas. The goal of the present experiments was to identify brain areas involved in the processing of suprasegmental auditory cues, as well as the respective time course of processing. An additional focus was placed on the effect attention (processing conditions) has on stimulus processing mechanisms.

***Experiment I*** focuses on rhythm in speech perception revealing an involvement of secondary cortical areas, namely the posterior superior temporal gyrus (pSTG), in rhythm perception. The involvement of the pSTG was differentially lateralized depending on the processing condition. Implicit processing revealed activity in left cortical areas. Explicit processing revealed activity in cortical areas of the right hemisphere indicating a top-down modulation on right secondary auditory areas. The study revealed furthermore, that the supplementary motor area (SMA), formerly associated with motor production is involved in the performance of a rhythm task indicating a role of the SMA in timing perception as well as production. ***Experiment II*** focuses on the neurophysiological correlate of speech melody (prosody) processing focusing on effects of processing conditions. Additionally, the study aimed to reveal the candidate region for interaction effects between semantic and prosodic processing observed in an earlier electrophysiological study. A neurofunctional interaction between prosodic and semantic processing was observed bilaterally in the posterior middle temporal gyrus (pMTG). Furthermore, results of the study revealed an effect of the processing mode reflected in frontal as well as temporal brain areas, namely, Brodmann area 44 (BA44), the posterior middle temporal gyrus (pMTG) and the planum temporale (PT). ***Experiment III*** investigated the electrophysiological correlate of auditory rhythm perception, focusing on differences between meter and rhythm perception as revealed in a specific

electrophysiological component (N1). Results revealed a differential effect the processing mode has on the N1. Rhythm processing revealed an N1 in attended and unattended processing. Meter processing revealed an N1 in attended processing only, suggesting different neurophysiological processes underlying the auditory perception of rhythm and meter.

To summarize, the present results indicate that temporal auditory perception, akin to melody perception, is processed in the posterior part of the superior temporal gyrus and that these neurofunctional mechanisms are affected by distinct processing conditions.

## Zusammenfassung

Thema der vorliegenden Arbeit sind die neuronalen Grundlagen suprasegmentaler auditorischer Signalverarbeitung am Beispiel von Sprache und Musik. Suprasgementale Signaleigenschaften ziehen sich über mehrere Silben und Wörter respektive über mehrere Töne und beinhalten insbesondere Melodie und Rhythmus.

Während die neuronalen Grundlagen der Melodieverarbeitung bereits relativ gut untersucht sind, gibt es noch wenig experimentelle Befunde zur Verarbeitung von akustischen Rhythmen, wobei hier insbesondere die Frage interessiert, ob die Verarbeitung von Rhythmus eine Lateralisierung auf eine Hemisphäre des Grosshirns aufweist. Neben der Identifikation von kortikalen Arealen zur Verarbeitung von suprasegmentalen Signalen liegt ein Fokus der Arbeit auf dem Einfluss von Aufmerksamkeit auf die neurofunktionellen Verarbeitungs-mechanismen.

Das ***Experiment I*** untersuchte die neuronalen Grundlagen von Sprach-rhythmusverarbeitung, welche im sekundären auditorischen Kortex, namentlich im posterioren Teil des gyrus temporalis superior (pSTG), identifiziert werden konnten. Interessanterweise zeigte sich eine von der Prozessierungsbedingung abhängige Lateralisierung. Explizite (aufmerksamkeitsabhängige) Sprachrhythmusverarbeitung war von einer signifikanten Aktivierung in der rechten Hemisphäre begleitet, während implizite (aufmerksamkeitsunabhängige) Verarbeitung mit einer signifikanten Aktivierung in der linken Hemisphäre assoziiert war. Überdies werfen die Ergebnisse dieser Studie ein neues Licht auf die Funktion einer bis anhin fast ausschliesslich mit motorischen Prozessen in Verbindung gebrachten Hirnregion, die area motoris supplementaris (SMA). Die Aktivierung der SMA in der Verarbeitung von Sprachrhythmus weist auf eine generelle Funktion dieser Hirnregion in zeitlicher Perzeption hin. ***Experiment II*** untersuchte die neuronalen Grundlagen von Sprachmelodieverarbeitung und insbesondere den Einfluss unterschiedlicher Aufmerksamkeitsbedingungen auf die Aktivität in involvierten Hirnregionen. Die Befunde zeigten einen Einfluss der Aufmerksamkeitsbedingung auf das Verhalten sekundärer auditorischer Areale, insbesondere das planum temporale (PT) und der posteriore Teil des gyrus temporalis medialis (MTG) bei der Stimulusverarbeitung. Abhängig von der Aufmerksamkeit wurden stärkere Aktivierungen bei der

Verarbeitung von kongruenten oder inkongruenten Stimuli beobachtet. Zudem zeigten sich Interaktionseffekte zwischen prosodischen und semantischen Stimuluscharakteristiken insbesondere im pMTG und dem gyrus frontalis inferior (BA44). ***Experiment III*** untersuchte die elektrophysiologischen Grundlagen von Rhythmusverarbeitung. Insbesondere wurde eine mögliche Unterscheidung zwischen den beiden rhythmischen Elementen Metrum und Rhythmus in Bezug auf eine Evozierung der N1 untersucht. Die Befunde zeigten einen Unterschied zwischen Metrum und Rhythmus, der sich im Zeitverlauf der Verarbeitung äusserte. Rhythmische Stimulusmanipulationen riefen sowohl unter expliziter als auch unter impliziter Verarbeitung eine N1 hervor, während metrische Stimulusmanipulationen nur bei expliziter Verarbeitung eine stärkere N1 im Vergleich zu nicht-manipulierten Tönen evozierten.

Insgesamt konnten die vorliegenden Untersuchungen die Verarbeitung suprasgementaler akustischer Signale in sekundären auditorischen Kortexarealen zeigen. Die Lateralisierung scheint jedoch mindestens im Falle der akustischen Rhythmusverarbeitung von der Aufmerksamkeitsbedingung beeinflusst zu sein. Die Verarbeitung von Sprachmelodie wird zudem von anderen linguistischen Charakteristiken beeinflusst, wie hier am Beispiel der Semantik gezeigt wurde.

# I Introduction

Auditory rhythm and melody perception is an integral part of the auditory experience. In music perception rhythm and melody constitute the two most important structural devices. Similarly, speech also relies on these two elements facilitating word identification and transporting linguistic as well as affective information (Friedrich, Kotz, Friederici, & Alter, 2004). The combining characteristic of melody and rhythm is their expansion over several auditory segments, which is the reason why they are referred to as *suprasegmental auditory signals* in auditory research.

The focus of the present thesis lies on the neural correlates of suprasegmental auditory signal processing, especially concentrating on rhythm processing. The brain areas involved in these processing mechanisms as well as their time course is investigated by means of functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). An additional central point of this work is the investigation of the influence attention has on these processing mechanisms. Whereas neuroimaging research has so far focused on the melodic aspects of suprasegmental speech processing, rhythm, more generally referred to as "timing", is only starting to gain more attention. However, language research has long since focused on speech rhythm trying to identify the exact phonological or acoustic correlate of speech rhythm. And this research is still ongoing. It is this lack of clarity in the definition that has probably prevented neuroimaging research from investigating the neural correlates of speech rhythm perception. In music, rhythm is more uniformly defined. However one of the challenging issues in that area is the question of a categorical difference between meter and rhythm, which could possibly be answered by neurofunctional research. Both, speech rhythm and rhythm processing in music are investigated in the present thesis.

The *second chapter* of this thesis depicts the central neurophysiological pathway of the auditory signal as well as the current anatomical and functional understanding of cortical areas involved in auditory processing. The *third chapter* then describes the nature of suprasegmental auditory processing and outlines current issues in research.

This is followed by three empirical studies in the form of independent manuscripts in the *fourth chapter*. In the *fifth chapter* the main results of these empirical studies are discussed in the context of the open questions posed in chapter two. The thesis closes with final remarks resulting from the work presented.

# II Auditory processing in the central nervous system

## The central auditory pathway

The path of the acoustic signal from the inner ear to the cortex includes several processing steps as indicated in figure 1. Here it is characterized focusing on two main principles namely frequency and amplitude encoding, and parallel processing.

In the cochlea the acoustic signal is transformed into a neural signal that then passes from the cochlea via the auditory nerve to the cochlear nucleus located in the medulla of the brain stem. The cochlear nucleus passes the signal to the other nuclei in the pons and midbrain, the superior olive, the lateral lemniscus and the inferior colliculus. The signal then advances to the medial geniculate complex (MGC) of the thalamus and subsequently to the primary auditory cortex, which in humans is Heschl's gyrus.

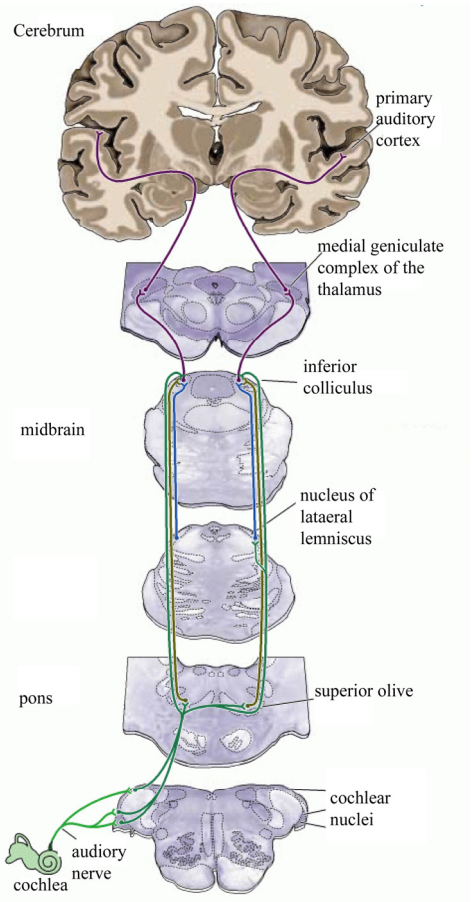



**Figure 1:** The central auditory pathway (Purves et al., 2001)

It is in the cochlea of the inner ear that the signal is frequency and amplitude encoded. Unlike other sensory systems, this is the only information entering the central auditory pathway. Spatial or topographic information, which in other sensory systems is an additional element of information, is encoded in the two devices here and subsequently retranslated into spatial information. Such processing requires an analysis in tremendous detail which most likely starts on a sub-cortical level (Krumbholz et al., 2005; Zatorre & Penhune, 2001; Warren & Griffiths, 2003).
In the cochlea the basis of the basilar membrane responds to high frequencies, whereas the apex of the basilar membrane responds to low frequencies. This processing structure called *tonotopy* is at least partially maintained throughout the central auditory pathway. Neural complexes, responsive to a specific frequency, known as isofrequency laminae have been found in tissue of the brain stem and thalamus (Takayanagi & Ojima, 2006), as well as the primary auditory cortex in both animal research (Nishimura, Shirasawa, Kaizo, & Song, 2007) as well as human research (Guiraud et al., 2007). Tonopy is lost in the secondary auditory cortex (Langers, Backes, & van Dijk, 2007).

Starting at the cochlear level, parallel processing continues through to the primary auditory cortex. But as the signal makes its way to the auditory cortex, integration of signals from both ears already takes place. In the brain stem, the superior olive as well as in the commissure of the inferior colliculi the neuronal signal is passed to the contralateral side, and thus information of the ipsilateral as well as of the contralateral side is integrated. In the primary auditory cortex, bilateral integration is reflected in the neural architecture. Neurons that display summation interactions with their bilateral neighbour are located next to each other, as are neurons displaying inhibition interactions. One of the main purposes of parallel processing on subcortical and cortical level is the spatial localization of sound (Brown, 2003). The horizontal integration processes mentioned above are completed by vertical integration mechanisms. The signal from the brain stem must pass through the ventral part of the medial geniculate complex (MGC) of the thalamus to enter the primary auditory cortex. However, functional connections from the cortex down to the medial part of the MGC also exist. Thus, it is evident that the interaction between primary auditory areas and thalamus is bi-directional, and that top-down integration

mechanisms already take place at this level (Schonwiesner, Krumbholz, Rubsamen, Fink, & von Cramon, 2007).

**Anatomic subdivisions of the auditory cortex**

Recent advances in the comprehension of auditory processing at the cortical level have been made through sophisticated techniques of tissue analysis as well as through functional fMRI research. Specifically, the anatomical parcellation of auditory cortices on the basis of cytoarchitectonic markers and recent advances in auditory functional fMRI research (see method section of this thesis) help to better understand auditory signal processing on a cortical level. The next section describes the current knowledge on the anatomy of the auditory cortex.

Primary and secondary auditory cortices are differentiated on a macro-anatomical basis (Fig 2a). In humans the *primary auditory* cortex is located bilaterally at the surface of the superior temporal plane and is called Heschl's gyrus. It is the rostral-most transverse gyrus and its course runs from the retro-insular region medially to the lateral surface of the first temporal gyrus (Steinmetz et al., 1989). The posterior sulcus of the Heschl's gyrus is called Heschl's sulcus. The *secondary auditory* cortices consist of additional transverse gyri located in the superior temporal plane. The planum polare anterior to the Heschl's gyrus (HG) is also considered to be part of the auditory cortex (Rademacher et al., 2001). One transverse gyrus lies anterior of the HG and posterior of the planum polare. One or more gyri lie behind the HG and are called the planum temporale (Leonard, Puranik, Kuldau, & Lombardino, 1998). The triangular planum temporale was first described by Galaburda (Galaburda & Sanides, 1980). Its borders have been described on the basis of macro-anatomical indices (Shapleske, Rossell, Woodruff, & David, 1999; Steinmetz et al., 1990) as being the superolateral margin of the STG, Heschl's sulcus, and the intersection of the posterior and anterior border at the medial side. While the posterior border has still not consistently been defined, it is most likely the ascending part of the posterior end of the Sylvian fissure.

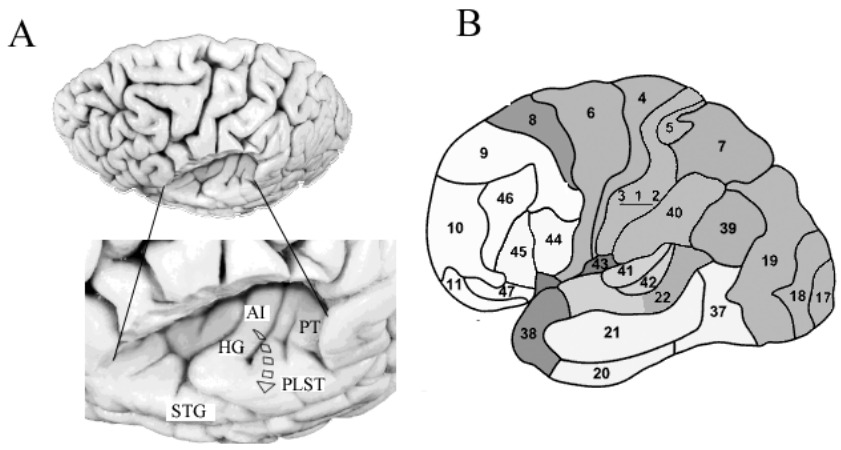



**Figure 2:** a) The superior temporal plane (Howard et al., 2000); b) Brodmann areas

The planum temporale displays a characteristic lateralization that has been measured using post-mortem volumetry (Shapleske et al., 1999; Westbury, Zatorre, & Evans, 1999) as well as in vivo magnetic resonance morphometry in infants (Preis, Jancke, Schmitz-Hillebrecht, & Steinmetz, 1999) and adults (Jancke & Steinmetz, 1993). Furthermore, there is evidence for a heritability of anatomical lateralization (Eckert et al., 2002). Additional *auditory association cortices* are the superior temporal gyrus and the superior temporal sulcus.

One of the earliest nomenclatures of the auditory cortex based on cytoarchitectonic characteristics was given by Brodmann who identified the areas BA 41 as the primary auditory cortex, BA 42 as secondary auditory cortex in the superior temporal plane and BA 22 as tertiary auditory cortex encroaching over the superior temporal gyrus and sulcus (Brodmann, 1909) (Fig 2b). In the 1980, another classical cytoarchitectonic analysis of the auditory cortex was undertaken (Galaburda et al., 1980). Authors performed post mortem analysis of primary and secondary auditory cortices in three human brains. Nissl staining revealed differences between primary and secondary auditory cortical areas and, furthermore, subdivisions of the temporal cortex were identified as secondary auditory cortex areas, namely, the planum temporale (PT) at the superior temporal plane caudal to Heschl's gyrus and the

parietal operculum (P0). This assumption no longer holds true for the parietal operculum.

Recently, new techniques have been developed to analyze the cytoarchitectonic structure of the auditory cortex. These now allow for the analysis of a greater number of individual anatomies (Morosan et al., 2001; Rademacher et al., 2001). By this means, three subdivisions of the primary auditory cortex, Te1.0, Te1.1, Te1.2, have been defined (Morosan et al., 2001). Approaches combining cytoarchitectonic with receptor-architectonic methods (Zilles et al., 2002) identified Te3, an area earlier summarized in BA22, as a separate area within clearly defined receptor-architectonic borders (Morosan, Schleicher, Amunts, & Zilles, 2005).

**Functional anatomy of the auditory cortex**

The auditory signal arrives from the thalamus and enters the cortex via Heschl's gyrus. It is well evidenced that tonotopy is still maintained in the *primary auditory cortex (Heschl's gyrus)* but ceases in the secondary auditory cortex. One study reported differences in tonotopy within the left and the right hemisphere with a less stringent structure in the left hemisphere (Ozaki & Hashimoto, 2007). Furthermore, electrophysiological research has investigated the time course of processing in the auditory cortex indicating serial processing along the medio-lateral axis starting from the most medial part of the supratemporal (Inui, Okamoto, Miki, Gunji, & Kakigi, 2006).

The *secondary auditory cortex* receives signals from the Heschl's gyrus by parallel processing streams (Inui et al., 2006) and seems not to be structured tonotopically anymore (Langers et al., 2007). Many studies have investigated the functional significance of secondary auditor cortex areas. A classical electrophysiological study suggested two functionally separate auditory pathways from investigations of the cortex of rhesus macaques. A ventral stream projects from the anterior part of the superior temporal gyrus to ventral prefrontal regions and is involved in object recognition. A dorsal stream specializing in auditory object localization projects form the posterior part of the superior temporal gyrus to the dorsal part of the prefrontal cortex (Romanski et al., 1999). Recently, an fMRI study has suggested a

comparable distinction in humans associating the posterior part of the STG with changing sound source location and the anterior part of the STG including Heschl's gyrus with pitch processing (Barrett & Hall, 2006).

The posterior part of the STG, specifically the *planum temporale,* has been the focus of interest due to its specific anatomic lateralization to the left hemisphere. However, the original hypothesis of a specific function of the planum temporale (PT) in language processing, suggested by the lateralization of speech processing to the left hemisphere and the anatomical lateralization of the PT, is now disregarded. Although both lateralization asymmetries are well evidenced, the two seem not to correlate (Josse, Mazoyer, Crivello, & Tzourio-Mazoyer, 2003; Jancke et al., 1993). Nevertheless, a shift in anatomical lateralization has been found to correlate with specific perceptive deficits as well as capacities. Dyslexics displayed a tendency for a smaller left planum temporale (Heiervang et al., 2000) and subjects with absolute pitch hearing display a stronger leftward asymmetry in anterior regions of the planum temporale as evidenced by voxel-based morphometric studies (Luders, Gaser, Jancke, & Schlaug, 2004; Keenan, Thangaraj, Halpern, & Schlaug, 2001). Moreover, subjects with developmental stuttering who profited from delayed auditory feedback displayed a rightward asymmetry in planum temporale, while subjects who did not profit from delayed feedback displayed a leftward asymmetry (Foundas et al., 2004).

More recent evidence suggests a domain-unspecific function of the PT in auditory processing. Griffith & Warren refer to the planum temporale as a "computational hub" involved in the analysis of auditory signals, specifically with respect to spectro-temporal patterns (Griffiths & Warren, 2002). Several studies have recently supported this assumption associatin the PT with pitch processing (Warren, Uppenkamp, Patterson, & Griffiths, 2003) and spectral (Griffiths et al., 2007). Conversely, with the dual stream model, the posteromedial part of the STG was found to be involved in spatial sound allocation (Warren et al., 2003; Zatorre, Bouffard, Ahad, & Belin, 2002; Altmann, Bledowski, Wibral, & Kaiser, 2007; Barrett & Hall, 2006).

The cortical areas anterior to the Heschl's gyrus, including the *planum polare,* are functionally less precisely characterized. The planum polare, in addition to the HG and PT, was found activated for melody perception in the contrast between melody and fixed pitch (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002). These authors suggested a hierarchy of pitch processing with melodic processing taking place anterolaterally to the primary auditory cortex. The anterior part of the superior temporal plane has often been associated with pitch perception, melody repetition and harmonization (Brown, Martinez, Hodges, Fox, & Parsons, 2004; Warren et al., 2003; Barrett & Hall, 2006) but also with the perception of acoustic "edges", that is, the offset of a tone (Herdener et al., 2007).

We can conclude that the more complex auditory signals are processed in a distributed network including secondary auditory cortex areas. While the functional specialization of cortical regions anterior to the Heschl's gyrus is still a matter of investigation, the posterior part of the superior temporal gyrus, often referred to as planum temporale, seems to be crucial for the analysis of complex auditory signals, specifically in association with spectral aspects of sounds. Lateralization abnormalities in heterogeneous symptom complexes such as dyslexia, developmental stuttering and absolute pitch perception further sustain the crucial involvement of this region in sound processing.

# III Current issues in auditory research: Suprasegmental auditory signal processing

The preceding chapter closed with the notion, that complex auditory signals are analyzed in a distributed network. For humans, speech and music are among the most complex auditory signals to be analyzed by the brain and each of these two fields has consequently been the subject of extensive investigations by neuropsychologists. The present thesis focuses on a specific aspect of these complex signals, namely suprasegmental auditory signals. In the following chapter the nature of suprasegmental auditory signals concentrating on the two main characteristics – rhythm and melody – will first be briefly introduced. Then the recent state-of-the art of neuroimaging research on suprasegmental processing is outlined, and open questions are introduced and related to the studies that are subsequently presented. This chapter will close with a summary of the issues investigated in the empirical part.

The terminus "suprasegmental" is used in linguistics and describes all speech characteristics, which extend over more than one sound segment, phoneme, syllable or word (Bussmann, 1990; Crystal, 1975). It includes pitch contour, intensity patterns, syllable length, and pauses in a verbal utterance. One of the most investigated suprasegmental element of speech is the pitch contour, which is also referred to as *speech melody* or *prosody*. A specific phonetic alphabet has been developed to describe the prosody of American English (Silverman et al., 1992; Hirst, 2005) and other languages such as German (Grice, Baumann, & Benzmüller, 2005). Prosody is very important in speech perception transporting not only linguistic information, as it is the case in statement or question intonation, but also emotional content. It can even influence the meaning of an utterance by conveying doubt or irony, for example. Suprasegmental speech elements, except for pitch contour, are described as *rhythmic or temporal* speech characteristics including syllable accents, or lengthening, pauses and speech tempo. Its importance in speech perception is undisputed, comprising online speech segmentation (Cutler, 1994; Jusczyk, 1999), lexical identification (Friedrich et al., 2004), and even language

acquisition (Nazzi & Ramus, 2003). Furthermore, practical experience shows that a strong foreign accent can render an utterance almost incomprehensible, even though syntax and semantic are perfectly correct.

Neurophysiological and neuroimaging research on suprasegmental auditory processing so far focused particularly on speech melody displaying a congruent picture of the brain areas involved. It appears that speech melody takes up special position within speech processing. While most speech processes are lateralized to the left hemisphere, in most circumstances speech prosody seems to be lateralized to the right in adults (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Friederici & Alter, 2004; Pell, 2005; Baum & Dwivedi, 2003) and even in young infants (Homae, Watanabe, Nakano, Asakawa, & Taga, 2006). These studies revealed activity in right superior temporal brain areas, which is completed with frontal activity under specific processing conditions (Hesling, Clement, Bordessoules, & Allard, 2005). However, the left hemisphere seems to play an important role as well, as indicated by impairment in prosody perception after agenesis of the corpus callosum (Paul, Lancker-Sidtis, Schieffer, Dietrich, & Brown, 2003).

Widely accepted models of auditory perception interpret these findings on the basis of a relative hemispheric specialization in the perception of acoustic characteristics. It has been proposed that the left hemisphere is specifically proficient in the temporal analysis of auditory stimuli while the right hemisphere is proficient in pitch and spectral analysis of auditory stimuli (Zatorre & Belin, 2001). Alternatively, lateralization effects are related to the specific temporal resolution in auditory perception. This model proposes that the left posterior auditory cortex is preferentially driven by rapidly changing acoustic cues, i.e. formant transitions, while right posterior auditory cortex is better suited to slowly changing acoustic cues, i.e. intonation contours (Poeppel, 2003). These two parameter-specific concepts are widely considered to be two sides of the same coin. Both concentrate their attention particularly to the posterior part of the superior temporal gyrus, namely the planum temporale (PT), as being critically involved in temporal segmentation processes (Griffiths et al., 2002).

*Experiment I - the hemispheric lateralization of speech rhythm:* The neural correlates of speech rhythm processing have not been investigated so far. Yet, this issue is of particular interest related to the question of functional lateralization, since rhythmic aspects of speech contain distinctive temporal characteristics. For the perception of these speech elements it is assumed that the analysis of speech characteristics requires the perception of both the short temporal window, as for the perception of syllable length, as well as the long temporal window, as rhythm perception is integrated over a whole sentence. Thus, findings on the perception of speech rhythm might further elucidate parameter-specific hemispheric lateralization in auditory perception. One reason for the hesitant manner in which neuroscience has approached speech rhythm, may be the difficulty encountered with the definition of the acoustic and linguistic correlates of speech rhythm. Speech rhythm is considered a conglomerate of different suprasegmental speech cues such as syllable duration, syllable stress, or pause (see introduction to manuscript I). The only study investigating rhythm perception in speech reported that damage in the left hemisphere effecting a clinical picture of non-fluent aphasia, led to a slower performance of a word grouping task (Aasland & Baum, 2003). Other lesions studies investigating auditory timing perception are incongruent, reporting left as well as right hemispheric involvement (Penhune, Zatorre, & Feindel, 1999; Di Pietro, Laganaro, Leemann, & Schnider, 2004). However, reproduction tasks seem to congruently rely on right hemispheric brain areas (Wilson, Pressing, & Wales, 2002; Riecker, Wildgruber, Dogil, Grodd, & Ackermann, 2002).

Experiment I thus focuses on the neural correlate of speech rhythm and meets the challenge of the rhythmic definition by an innovative approach exaggerating the typical speech rhythm of German in order to identify cortical areas necessary for its perception. Furthermore, the study specifically tests the hypothesis of a right hemispheric involvement in speech rhythm perception.

*Experiment II – Do we trust prosody when in doubt?* Irony is one example when prosody changes the meaning of an utterance. Most listeners would agree that the prosodic content of a sentence is just as valid as its semantic content and that prosody can even overrule semanticity.

Although prosodic processing has been investigated quite extensively, interaction processes between suprasegmental processing and the processing of other aspects of speech are yet less investigated. Since the interaction between sentence level prosody and syntax is most prominent, as for example in the understanding of question vs. statement intonation, some studies have focused on this interaction. Humphries et al have identified an interaction between prosody and syntax in the anterior temporal cortex bilaterally along the superior temporal gyrus specifically in the left anterior superior temporal cortex (Humphries, Love, Swinney, & Hickok, 2005). ERP studies have revealed interaction effects as well (Eckstein & Friederici, 2005; Eckstein & Friederici, 2006; Hahne, Eckstein, & Friederici, 2004). Although most obvious, interaction effects between semanticity and prosody have only been investigated in two ERP studies (Kotz & Paulmann, 2007; Astesano, Besson, & Alter, 2004). It is suggested that an interaction effect appears approximately 800 ms after stimulus perception.

Experiment II thus focuses on speech melody processing and the neurofunctional correlate of its interaction with semantic processing.

*Experiment III - Timing in music perception:* Finally, suprasegmental auditory characteristics are not only present in speech but also in other auditory systems, namely in music. Functional parallels between music and language as a means of communication have been drawn by scientists for over a century (Darwin, 1871). Recent literature has followed Darwinian reasoning described music as a "protolanguage" serving communicative objectives (Merker, 2000). Also structural parallels are numerous, among them complexity, generativity, hierarchical structure, and emotionality.

Among the first cognitive researchers suggesting that the two systems, music and language, rely on similar "mind systems" was Lerdahl who specifically focused on similarities concerning the temporal structure of the two systems (Lerdahl, 2001). He followed a purely descriptive approach analyzing a poem on the basis of musical rhythm taxonomy. Other descriptive approaches have reported culture-specific similarities in the rhythm of music and language for French and German (Patel, 2003; Patel & Daniele, 2003). The empirical observation that amusic subjects performed as badly in the discrimination of speech material as in the discrimination

of music material, might support this assumption (Patel, Peretz, Tramo, & Labreque, 1998).

However, specifically in the temporal domain differences between music and language are to be found. For example, whereas music is often metrical, speech is only metrical in some circumstances (Fitch, 2006). However, one can argue that meter in speech can be present in certain circumstances, such as in poetry. And some languages, German in particular, have a strong tendency to isochrony and are thus more metrical than others. Similarly one can argue that music is not radically metrical either. In fact if it were, it would not be perceived as "musical".

With respect to the temporal resolution, the timing in both systems is comparable. The timing of tones in music lies approximately in the same time range as phonemes and syllables. And timbre has a temporal resolution that is comparable to voice-onset-time in speech. Finally, the temporal phrasing of a musical piece has durations of mostly more than a second, which is about the time duration of a sentence.

Newer models of hemispheric lateralization models have explicitly set the ground for comparisons between speech and musical signals. They moved away from domain specific lateralization focusing on the acoustic aspects of auditory signals, an approach which as been followed on the segmental level of acoustic processing in particular (Zaehle, Wustenberg, Meyer, & Jancke, 2004). On the suprasegmental level, this approach has so far not been followed, neither for melody nor for rhythm processing.

One important question to ask in musical rhythm research is whether there is a possible differentiation between rhythm and meter (for a detailed description, see the introduction to experiment III). Behvioral research has proposed a categorical difference between the two rhythmic characteristics (Peretz & Zatorre, 2005). However, the corresponding neuropsychological evidence is still lacking.

Thus, experiment III aimed to investigate the neurophysiological correlate of auditory rhythm and meter perception. As a means of investigation, EEG was chosen as it is specifically used to isolate possible differences in the early time window after stimulus perception.

*Paradigmatic approach of the three studies:* The three experiments reported in this thesis are tied together not only by their topic of investigation, but also by a specific paradigmatic approach. Rather than being restricted to only one processing context, the studies investigate the neurophysiological mechanisms of interest in two different processing contexts by varying how much attention subjects had to pay to the specific stimulus characteristic. In one condition, the stimulus of interest is processed explicitly (attentional processing condition) whereas in the other processing condition the stimulus is processed implicitly as subjects focus their attention on another acoustic characteristic of the stimulus material (unattentional processing condition).

The observation that the involvement of frontal brain areas is task dependent is generally accepted (Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006; Noesselt, Shah, & Jancke, 2003; Binder, Liebenthal, Possing, Medler, & Ward, 2004; Hsieh, Gandour, Wong, & Hutchins, 2001). However, there is a growing body of evidence suggesting that top-down mechanisms also significantly influence activity in primary and secondary sensory areas (Binder et al., 2004; Noesselt et al., 2003; Davis & Johnsrude, 2003; Jancke, Mirzazade, & Shah, 1999). That is, depending on the context of stimulus perception or on the task performed on these stimuli, activity in temporal brain areas is modulated. In some circumstances the brain areas significantly involved are even altered, as indicated by task dependent functional lateralization effects (Brechmann & Scheich, 2005; Scheich, Brechmann, Brosch, Budinger, & OHl, 2007; Tervaniemi & Hugdahl, 2003). Despite this observation, most researchers nevertheless only use one processing condition manipulating the characteristics of the stimulus material. This is most likely due to the difficulties two processing conditions pose on the experimental paradigm. Among them are the substantially longer duration of the experiment, the problems between-subject comparisons pose on fMRI analysis, as well as the loss of power in between-subject fMRI models due to the higher number of experimental factors (Poline, 2003). However, progress in fMRI-research is dependent on the comparison of a great number of studies. The fact that different studies often vary in the processing condition of a specific stimulus characteristic makes it very difficult to draw a comparison. Therefore, all three experiments of the present thesis include two

different processing conditions in addition to the manipulation of stimulus characteristics.

**Empirical studies**

The general questions investigated in the empirical part were the the neural correlates of suprasegmental auditory processing, that is, rhythm and melody perception, focussing thereby on questions of functional lateralization. Furthermore, the various experimental designs were aimed to separate brain areas involved in attentional and unattentional processing of stimulus characteristics, hereby trying to resolve the question of modulatory effects the processing conditions have on the brain areas involved in specific perceptual processes. This approach led to the following questions:

*Experiment 1:* The goal of the first experiment was to identify brain structures involved in the processing of speech rhythm. More precisely, the study first aimed to identify cortical brain structures involved in the performance of a speech rhythm task. Second, the study addressed the question of cortical regions related to speech rhythm perception in two different processing conditions, namely explicit and implicit processing.

*Experiment 2:* The goal of the second study aimed at identifying brain structures involved in the processing of speech melody. This issue has already been investigated using several different approaches. This study focused on the issue of differential context dependent processing. A second aim was to investigate the interaction between prosodic and semantic processing in their relative associated brain areas.

*Experiment 3:* The goal of the third study was to identify the electrophysiological correlate of auditory rhythm processing profiting from the high temporal resolution of EEG methods. More precisely, differences in the correlates of meter and rhythm processing, the two main structuring categories of temporal organization, were investigated. Additionally, the effect explicit and implicit processing has on these early correlates was investigated.

# IV Methods

The empirical part of this thesis refers to two neurophysiological methods used to investigate cognitive processes in the brain, namely functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). The two methods rely on different correlates of neural activity. fMRI relies on activity induced blood perfusion changes and allows the identification of brain areas involved in perceptive or cognitive processes with a fairly high spatial resolution (mm). EEG relies on activity induced electrical potentials derived from the scalp surface and measures with a high temporal resolution (ms). The following section provides the reader with the necessary knowledge to understand the two methods utilized in the empirical section[1].

## Functional Magnetic Resonance Imaging

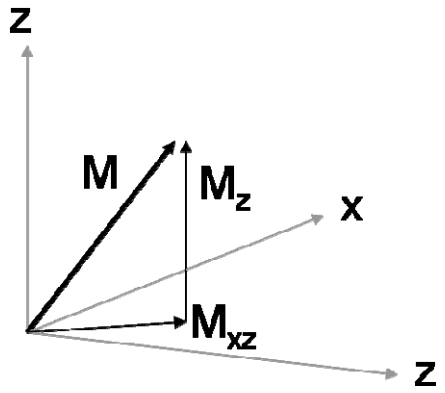
*Basic Principles of Magnetic Resonance Imaging (MRI)*

The physical basis of magnetic resonance imaging is the nuclear magnetic resonance of hydrogen atoms in organic tissue. It relies on the specific physical behavior of nuclear protons. Nuclear protons comprise one positive charge and rotate with a characteristic angular momentum (spin). The rotating electrical charge elicits a magnetic momentum ($\vec{\mu}$), which then behaves like a magnetic dipole in a magnetic field (B). When placed into a steady magnetic field, spins orient either in a parallel or anti-parallel orientation to the magnetic field adopting a corresponded precession known as the larmor frequency. Parallel and anti-parallel orientations differ in their energetic level, with the parallel orientation being energetically more favorable.

$$\omega_0 = \gamma B_0$$

As indicated above, the larmor frequency is directly proportional to the strength of the magnetic field in Tesla [T] and to a constant $\gamma$ that is characteristic for $H^+$-protons. Aligned protons induce a longitudinal magnetization ($M_z$) in the direction of the magnetic field ($B_0$).

To derive a signal, a short-lasting high frequency pulse (HF) chosen with $\omega_0$ to alter the spin energy is induced. This alteration leads to a change in longitudinal magnetization rotating in the xy-plane in the direction perpendicular to the longitudinal field (transversal magnetization, $M_{xy}$) with the frequency $\omega_0$. This change in magnetization is measured via an induced voltage in a receiver coil.




F**igure 3:** Magnetization components longitudinal magnetization (**$M_z$**), transversal magnetization (**$M_{xy}$**), and resulting net magnetization (**M**).

After deflection protons slowly regain their original (energetically optimal) orientation. The relaxation time necessary to regain the energetically optimal state differs depending on the local environment in which a proton is located, thus, on the tissue characteristics. The relaxation time of a specific tissue is characterized by the parameters T1 and T2. T1 is a constant and is defined as the time necessary to regain 63% of the longitudinal magnetization. It depends exponentially on Bo. T2 is the time after the transversal magnetization decreased on 37 % of its initial value. It describes the process of energy exchange between protons and is tissue specific.

The 3-dimensional space is depicted in the form of voxels (e.g. 3x3x5 mm) and measures on each voxel are written into a matrix file. Three dimentional measurements are achieved by inducing two different linear gradients into Bo, leading to variable T1-values. A short duration gradient pulse measures the third spatial dimension. The signals measured are decomposed in a 3-dimensional Fourier analysis.

[1] The following chapter refers to Frahm et al. (Frahm, Fransson, & Krüger, 1999)

*Functional Magnetic Resonance Imaging (fMRI)*

Functional MRI allows the mapping of specific cortical activity by measuring blood-perfusion changes induced by neural activity. This correlate is referred to as a blood-oxygen-level-dependent (BOLD) signal change (Buxton, Wong, & Frank, 1998). The BOLD signal shows a characteristic decrease in signal intensity (initial dip) followed by a signal intensity rise with a maximum after 5-8 seconds. After 10-16 seconds the baseline intensity is regained (Fig. 4).

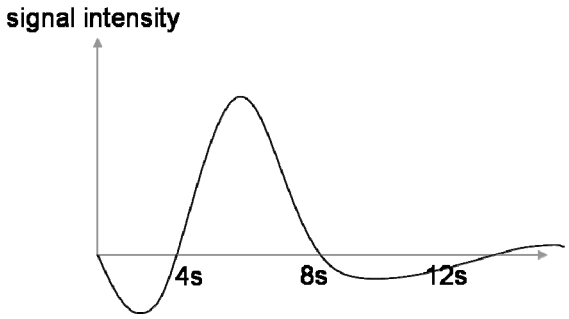



**Figure 4:** The time-course of the BOLD signal

Neural activity causes a rise in cerebral blood flow, which leads to a change of the local ratio between oxyhemoglobin and deoxyhemoglobin (hemodynamic response). After an initial decrease of oxyhemoglobin due to the consumption of oxygen (initial dip) there is an increase in blood flow and therefore an increase in oxyhemoglobin observed. The peak of this increase is reached 5-8 seconds after the stimulus onset that is assumed to cause the neural activity and returns to baseline after 10-16 seconds (Fox & Raichle, 1986). It is that change in the ratio of deoxyhemoglobin to oxyhemoglobin that is measured by fMRI techniques**.**

Blood hemoglobin has different magnetic characteristics depending on its state of oxygenation. While oxygenated hemoglobin has only a small magnetic moment and is diamagnetic, deoxygenated hemoglobin is paramagnetic. That is, deoxygenated hemoglobin behaves like a small magnet due to the iron in its molecules. The paramagnetic property of deoxyhemoglobin was originally detected by Ogawa et al. (Ogawa, Lee, Kay, & Tank, 1990).

*fMRI Data Analysis*

In this section the general steps of neuroimaging data analysis are detailed. Firstly, the three basic pre-processing steps are explained (Josephs, Turner, & Friston, 1997). This is followed by a description of ER-fMRI data analysis including model estimation and statistical testing by implementing the general linear model (GLM) according to Friston (Friston, 2004).

*Pre-processing* algorithms are applied in order to improve the data quality and to remove artifacts, i.e., non-task related variance. This form of correction is often referred to as the improvement of the signal-to-noise ratio (SNR) and consists of a number of steps. In the motion correction step, misalignment, which occurs as a result of the subjects moving during an experimental session, is corrected. This is achieved by aligning the functional volumes in space by means of rigid-body transformations. The software program SPM employs a least-squares approach and a 6 parameter (rigid body) spatial translation. In this approach, one image in a time-series serves as a reference scan against which all other scans are aligned. Motion correction has several limitations, including artifact-related limitations, which can result in a loss of data around the edges of the image volume. Temporal variance, which occurs as a result of temporal differences during acquisitions, is corrected for in a second step. This slice time correction realigns all the slices of one acquisition using a linear or sinc interpolation.

Functional images are usually collected to perform a subsequent analysis of random-effects. Therefore, the third preprocessing step spatially (stereotactically) normalizes images into a standardized space that is defined by an ideal model or template image (Ashburner & Fristson, 1999). In this case, the Montreal Neurological Institute reference coordinate system for the standard brain (a combination of 152 MRI scans on normal, right-handed subjects) was used.

Baseline drifts in fMRI time are low-frequency components of the signal that are due to multiple physiological and technical factors. They are generally much slower than the frequency of the signal of interest, which is correlated to the stimuli and the cognitive paradigm and can be removed, either in the preprocessing stage or during

model estimation, by applying a fast Fourier analysis and a subsequent high-pass filter. Cut off frequencies are usually chosen not higher than half the frequency of the signal of interest. A high-pass filter lets high frequencies pass and removes low frequencies. As a fourth optional processing step, a Gauss filter can be applied to spatially smooth the data.

The *statistical analysis* of the data is based on a general linear model (GLM), which is applied voxel-by-voxel. The general linear model relies on the linear multiple regression being extended in such a way that all scalar levels can be calculated. This model characterizes the relationship between the experimental factors and the observed data. In the design matrix the experimental regressors used to specify the model are convolved with a hypothetical bold response. Model estimation results in parameter estimates (Betas) describing the relationship between the observed data and the specific regressors defined in the model. It is these parameter estimates that are then used to statistically test the hypothesis.

Thus, the first step of a statistical analysis requires that a model containing condition-specific covariates be constructed. These models can be specified in different ways, and we distinguish between parametric and non-parametric models. In parametric models an assumption is made about the relative differences in brain activation between conditions. Neuropsychological experiments often define non-parametric models that are able to capture any difference between the experimental conditions (Poline, 03 S. 5). For this purpose a separate covariate is defined for every experimental condition containing the values 1 or 0 where 1 defines the onset of a given experimental condition. However, non-parametric models do not allow any specific form of increase or decrease (e.g. linear or quadratic) between conditions in the data to be tested. Nevertheless, the shape of an increase can be inspected post-hoc by displaying the parameter estimates (Poline, 03, S. 6). Factors that influence the signal but are of no interest, such as slow frequency drifts or environmental factors, are included in the model as regressors.
It is important to know that there is not only one way to specify a model for a given experiment. From a statistical point of view, the best model is the model that explains the variance in the data best, thus resulting in the least error variance.

However, one has to consider that a model to be defined has to test the specific hypothesis. Thus, from this point of view a good model meets the criteria of the hypothesis to be tested. It aims to answer the given hypothesis and excludes additional variance if necessary. In the case where too many questions need to be answered in one model, this often leads to multiple hypothesis testing at any one given voxel. This consequently increases the risk of false positives (Poline, 2003, S. 3).

Finally, the regressors or covariates are convolved with a hypothetical BOLD response in order to best fit the observed data, resulting in a *design matrix*. The convolution with a hypothetical bold response usually combines several basis functions called "temporal basis functions". Most often a set of gamma functions is used, since the shape of the Gamma function is similar to the BOLD response. A Fourier set or an "informed" basis function set, containing one actually measured BOLD response in the set, is also used. After the model definition, the *estimation* of a linear regression of the form $y = x_1\beta + x_2\beta + x_3\beta + ..... \times \varepsilon$ is performed for every voxel-time series ($y$). The regressors of the design matrix are linearly combined ($x_1, x_2, x_3, x_4$) encoding the model. Their relative quota in the signal measured for all time steps are expressed in vector $\beta$. The higher the value of Beta the more variance in the signal can be explained by this specific regressor. The residual variance in the signal is encoded by the vector $\varepsilon$. The error is assumed to be normally distributed and uncorrelated between voxels. The solution of this linear regression represents the best fitting equation that explains the values measured at each time step on a specific voxel. It is calculated in such a way that the sum of squared errors is minimized (least square estimation).

*Statistical inferences* are made by coding specific contrasts. In this way, the model compares the actual voxel-by-voxel time series of the signal to the hypothesis for the measured functional data, and so tests the hypothesis. Then, the beta-coefficients obtained for every single voxel in the estimation of the linear regression are used to contrast the different experimental conditions. The following section explains how hypotheses are tested in SPM using contrast vectors.

A contrast is essentially a linear combination of parameter estimates ($c' \times \beta$) and defines which conditions are to be compared. For example, the contrast vector c' = (1 0 0 -1 0) contrasts the first experimental condition with the fourth experimental condition. It is important to notice that the contrast of estimated parameters is essentially a sum of parameter estimates. If c' = 1 0 1 -1 -1 then the contrast of the estimated parameters computes $1 \times \beta_1 + 0 \times \beta_2 + 1 \times \beta_3 + 1 \times \beta_4 + 1 \times \beta_5$ .

The statistical significance of an observed effect can be tested using a t-test or an F-test. In SPM both tests are calculated for every voxel resulting in a matrix of t- or F-values. A t-test is used to test a one-dimensional contrast. The null hypothesis thus assumes there is no significant difference in the signal between two conditions. Its test statistic is the contrast of the estimated parameters divided by the square root of the variance estimate.

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

In other words, it is the effect size divided by the uncertainty of effect. It is important to notice that the variance estimate depends on 1) noise, that is, the residual variance in the signal, 2) the number of measurements, 3) the correlation between the measurements, and 4) the dependency of the regressors in the x-matrix. Thus, the more variance we find in the signal of a voxel, the less sensitive the signal in this voxel is for the experimental model.

The *second-level* analysis refers to the group statistic aiming to obtain multi-subject comparisons. Usually, a one-sample t-test is performed in order to test whether the mean contrast across a group of subjects is different from zero at a given voxel.

In order to reliably identify brain regions that exhibit differences in activation, a significance threshold is applied. For example, the frequently-used threshold of $p<.05$ corrected or $p<0.001$ uncorrected for multiple comparisons is applied.

*The mapping between structure and function*

Neuroimaging data collected to conduct a random-effects analysis are generally normalized to a standard brain such as the Talairach brain (Talairach & Tournoux,

1988), the MNI-brain which is an average of 152 normal MRI sans and which is the standard template in SPM. Functional images are usually visualized on an individual brain (Colin brain) that is normalized to the standard template. Anatomical mapping of functional activity to anatomical locations has long been done on the basis of macroanatomical landmarks which are also implemented in automatic anatomic labelling tools such as the AAL (Automatic anatomical labelling (Tzourio-Mazoyer et al., 2002). Recently, probability maps that account for individual variability in anatomical structuring have been created. They are either based on anatomical landmarks (Westbury et al., 1999) or on cytoarchitectonic and receptor distributive mapping (Rademacher et al., 2001; Morosan et al., 2001). The latter have subsequently been implemented in an automatic labelling tool (Eickhoff et al., 2005).

*Specific acquisition paradigms in auditory fMRI*

A specific problem faced by auditory fMRI research in the acquisition of data is the acoustic noise produced during acquisition. The bulk of this noise is produced mechanically by the Lorentz forces which result during the rapid alterations of currents within the gradient coils and in the presence of a strong static magnetic field (Mansfield, Glover, & Beaumont, 1998). For other noise sources see Moelker & Pattynama (Moelker & Pattynama, 2003). The noise-level depends, amongst others, on the repetition time (TR), field of view (FOV) and echo time (for a review see McJury & Shellock (McJury & Shellock, 2000). fMRI typically uses echoplanar (EPI) sequences to collect a complete image in one radiofrequency (RF) excitation of the spin system. These sequences involve fast switching gradients that produce particularly high amplitude noises of 115-130 dB SPL (Hattori, Fukatsu, & Ishigaki, 2007; Shellock, Ziarati, Atkinson, & Chen, 1998; Foster, Hall, Summerfield, Palmer, & Bowtell, 2000).

The consequence of this noise is twofold. Firstly, fine grained differences between auditory stimulus conditions are virtually unintelligible due to the acoustic noise and psychoacoustic effects might alter the perceived stimulus pitch and intensity. Secondly, it is assumed that the scanner noise interacts with the auditory stimulus processing on a physiological as well as on a cognitive level. Physiologically, saturation effects in the primary and secondary auditory cortex are assumed, which

lead to a lower sensitivity of the BOLD response to experimental stimuli in these cortical areas. Cognitively, enhanced attention due to difficulties in intelligibility modulates activity in primary and second auditory cortex and leads to additional attention related activity in frontal cortical brain areas (Shah, Jancke, Grosse-Ruyken, & Muller-Gartner, 1999; Belin, Zatorre, Hoge, Evans, & Pike, 1999; Scheich et al., 1998; Rinne et al., 2007).

In order to improve the brain's sensitivity to auditory stimulation in fMRI, several solutions have been proposed. Passive solutions are the use of ear plugs, which is standard in MRI scanning. However, this leads to unequal noise attenuation over the hearing range since high-frequency sounds are more attenuated than low-frequency sounds. Manipulations of scanning sequences have been suggested decreasing the level of gradient pulsing in an imaging sequence. However, this often leads to decreased signal-to-noise ratio (SNR) (Cremillieux, Wheeler-Kingshott, Briguet, & Doran, 1997; McJury et al., 2000). Active noise cancellation (ANC) techniques have been tested by introducing a sound which is the inverse of the original noise (Chen, Chiueh, & Chen, 1999).

The solution favoured in the present empirical studies follows the principle of sparse and clustered-sparse acquisition (see Experiment I, Fig. 1). The rational of these procedures is the presentation of stimuli in silence followed by the acquisition of single volumes of functional images close to the expected maximum of the BOLD response, controlling for perceptual as well as physiological confounds in auditory cortical areas related to the scanner noise (Hall et al., 1999; Belin et al., 1999; Shah et al., 1999). Specifically in experiments aiming to investigate primary and secondary cortical areas it appears that a clustered temporal acquisition is advantageous displaying a stronger functional response along the superior temporal plane in particular (Schmidt et al., 2007). However, one has to consider that the analysis of data form the clustered and sparsed temporal acquisitions has specific restrictions, amongst others that only the box-car or the frequently used FIR functions can be used to model the BOLD response and it is thus, not possible to analyse the time course of the BOLD response using these acquisition designs.

**Electrophysiological Measurements of brain activity**

The electroencephalogram measures neural activity of the cortex on the scalp surface. Hans Berger (1929) was the first to measure spontaneous EEG on the human scalp. Spontaneous EEG is mostly used in clinical research. In neuropsychological research one takes advantages of the possibility of measuring event-related potentials (ERPs) with a fairly high temporal resolution up to sub-milliseconds range. The following section describes the electrophysiological basis of the cortical potentials measured in EEG and describes the procedure of ERP analysis.

*Electrophysiological generators of scalp potentials*

The physiological generators of the electrical scalp potentials measured are the postsynaptic potentials of pyramidal neurons in the cortex. A cortical neuron has a typical resting potential of -70 μV with a more negative charge in the interior of the neuron. Neural activity is accompanied by a negative or positive shift of this resting potential leading to hyper-polarization or depolarization, respectively. A depolarization is induced by excitatory neurons and facilitates the generation of an action potential. These shifts in postsynaptic potentials induce an electrical dipole. Electrical dipoles are derived on the cortical surface under two preconditions, namely that the synchronous activity of about 50'000 neurons is necessary, and that only neurons aligned perpendicular to the cortical surface, contribute to the cortical potential. These are mostly pyramidal cells with their cell body in layer V and dendrites in layers I and II of the cortex. Moreover, the relative position of a neuron to other neurons influences the distant effect. Newer findings even indicate that the arrangement of Na+ and K-channels influence the distant effect of a cell (Murakami & Okada, 2006).

Spontaneous EEG contains potentials shifts of different frequencies that are depicted by letters of the Greek alphabet. Alpha-activity for example typically displays an oscillation of 8-13 Hz, amplitude of 20-120 μV, and is the dominating activity in most EEGs. Statistical analysis of theses frequency bands is done on the basis of energy. Spontaneous activity is induced by rhythmical activity probably generated by thalamo-cortical as well as cortico-cortical neuron complexes.

*Event-related potentials*

The first event-related potential on the human brain was described for auditory stimulation (Davis, 1939). Event-related potentials display a lower frequency than the frequency bands of spontaneous EEG and appear time-locked to the occurrence of a stimulus. Their identification was classically done by the polarity, latency and scalp distribution. More recent theoretical literature, however, focuses on the importance of the neuroanatomical generator site and cognitive function for defining an ERP (Luck, 2005). For the purpose of this study we focused on exogenous auditory components that depend on the auditory stimulus characteristic as well as on the vigilance of the subject. These auditory components are derived in a short, middle and long latency window (Fig. 6). Here of particular relevance are potentials with a long latency, that is, in a latency window of 100-300 ms. More specifically, it is the N1 originating from the primary and secondary auditory cortex that is derived. This component is sometimes accompanied by a positive shift 50 ms post stimulus onset and depends on selective attention (Hillyard, Hink, Schwent, & Picton, 1973).

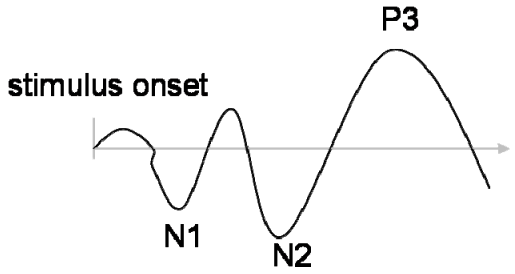



**Figure 6:** ERP components to auditory stimuli

*Recording and analysis of human scalp EEG*

The EEG is recorded on 32 or 64 electrodes positioned according to a standardized system, the 10/20-system with respect to the nasion, inion and left and right pre-auricular points (Jaspers, 1958). In the EEG-study of the present work, a modification of this standard system, in the form of the 10/10 system with 32 electrodes was used (Fig. 7). Here the distances between the electrodes average ten percent distance between the respective landmarks.

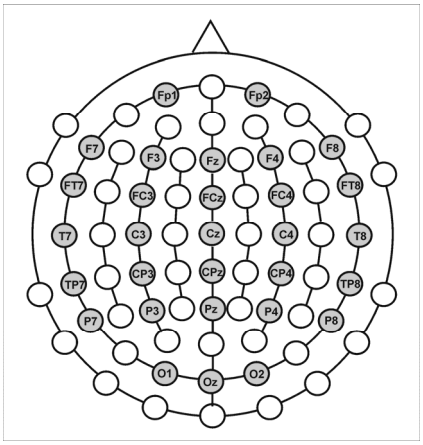



**Figure 7:** Locations of 30 electrodes following the 10/10 system.

Critical parameters in EEG recording are the sampling rate, which is the Nyquist frequency. It should be at least twice the fastest frequency present in the measured signal. Potentials are measured in reference to either a physical reference electrode (earlobes, nose or mastoids) or an average reference consisting of an average potential over all electrodes. Two pre-processing steps, artifact correction (Jung et al., 2000) and filtering, are performed in order to remove unspecific variance of the data. A baseline-correction is sometimes performed in within-subject designs to correct for low-frequency drifts in the data. Then experimental trials are averaged per subject. An ERP-random-effects analysis is subsequently performed on the peak amplitude and latency of a specific component.

# V Empirical part

## Experiment I: The neural correlate of speech rhythm as evidenced by metrical speech processing

Authors: Geiser E., Zaehle, T., Jancke, L., & Meyer M.
Pubished in Journal of Cognitive Neuroscience

**Abstract**

The present study investigates the neural correlates of rhythm processing in speech perception. German pseudo-sentences spoken with an exaggerated (isochronous) or a conversational (non-isochronous) rhythm were compared in an auditory functional magnetic resonance imaging (fMRI) experiment. The subjects had to perform either a rhythm task (explicit rhythm processing) or a prosody task (implicit rhythm processing). The study revealed bilateral activation in the supplementary motor area (SMA) extending into the cingulate gyrus and in the insulae extending into the right basal ganglia (neostriatum), as well as activity in the right inferior frontal gyrus (IFG) related to the performance of the rhythm task. A direct contrast between isochronous and non-isochronous sentences revealed differences in lateralization of activation for isochronous processing as a function of the explicit and implicit task. Explicit processing revealed activation in the right posterior superior temporal gyrus (pSTG), the right supramarginal gyrus, and the right parietal operculum. Implicit processing showed activation in the left supramarginal gyrus, the left posterior STG, and the left parietal operculum. The present results indicate a function of SMA and insula beyond motor timing and speak for a role of these brain areas in the perception of acoustically marked temporal intervals. Secondly, the data speak for a specific task-related function of the right IFG in the processing of accent patterns. Finally, the data sustain the assumption that the right secondary auditory cortex is involved in the explicit perception of auditory suprasegmental cues and moreover, that activity in the right secondary auditory cortex can be modulated by top-down processing mechanisms.

**Introduction**

The flow of speech is commonly perceived as having a particular rhythmic pattern. This rhythm is a pivotal structuring element of speech which is crucially involved in segmentation processes subserving language comprehension and even acquisition.

Recent evidence for this crucial function of speech rhythm has been found in electrophysiological research. Speech segmentation or the parsing of a sentence into words has been shown to rely on global patterns of prosodic phrasing which includes rhythmic and durational grouping as well as tonal pitch (Frazier, Carlson, & Clifton Jr., 2006). At the word level, research has demonstrated that rhythmic patterning of German utterances facilitates spoken word identification as evidenced by an electrophysiological priming study (Friedrich, Kotz, Friederici, & Gunter, 2004). Depending on the rhythmic organization of one's native language different segmentation strategies are observed. For French-speaking adults the syllable appears to be the unit of segmentation (Segui, Djupoux, & Mehler, 1990), whereas English-speaking adults are guided by information about typical word-stress patterns (Cutler & Norris, 1988). With respect to the segmentation strategy, the German language is comparable to the English language. The view that the rhythmic properties of language shape listeners' speech processing strategies has led to hypotheses of how infants develop efficient speech segmentation procedures in the acquisition of their mother tongue. There is evidence that infants' sensitivity to linguistic rhythm allows them to discriminate (non-native) languages from different rhythmic classes (Nazzi & Ramus, 2003) suggesting a pivotal role of rhythm in language acquisition. Similar discrimination capabilities have even been reported from primate studies investigating tamarins (Tincoff et al., 2005).

The intuitive notion that spoken languages have characteristic underlying rhythmic patterns has resulted in a considerable amount of linguistic research aimed at assessing the exact acoustic characteristics of speech rhythm. This research was driven by the hypothesis of isochrony as an underlying principle of linguistic speech rhythm. Thus, a distinction is often made between stress-timed languages in which periodicity is assumed to be based on the regular occurrence of stressed syllables (e.g. English, German, Dutch) and syllable-timed languages such as French, Italian,

or Spanish in which regularity is taken to be the outcome of all syllables being roughly of the same length (Pike, 1945). However, the phonological evidence for isochrony as the organizational principle of linguistic rhythm is still not unequivocal. As has been shown, the duration of interstress intervals in English is proportional to the number of syllables they contain, but can vary in duration influenced by the specific types of syllables they are comprised of as well as by the position of the interval within the utterance (Bolinger, 1965; Lehiste, 1977). Moreover, new statistical analyses of the speech signal, for instance of consonantal and vocalic intervals (Ramus, Nespor, & Mehler, 1999) or of the auditory prominence of phonetic events (Lee & Todd, 2004) have also been inconclusive with respect to the identification of the acoustic characteristics of speech rhythm. Rhythmic differences between languages, context dependent variability of speech rhythm and inter-individual differences make identifying the acoustic as well as the phonological elements constituting speech rhythm experimentally difficult. Whereas linguists classify speech rhythm together with speech melody as prosodic aspects of speech using the term unspecifically for suprasegmental speech characteristics extending over more than one sound segment or phoneme, psycholinguistic researchers investigate speech rhythm as a concept in its own right. Such studies have included the field of language acquisition (Nazzi & Ramus, 2003), subsequent language learning (Curtin, 2005), speech segmentation (McQueen, Otake, & Cutler, 2001) and speech typologies (Ramus, Nespor, & Mehler, 1999).

With respect to the neural correlate of speech rhythm reliable evidence is still sparse**.** Since speech rhythm must be defined as a conglomerate of parameters with suprasegmental cues such as syllable duration, syllable stress, or pause being most important, it comes as no surprise that neuropsychological research has focussed initially on these phonetic cues. In particular, specific event-related brain potentials, the "closure positive shift" (CPS) or the P350, have been observed to be modulated by the processing of speech accent or speech pauses (Steinhauer, Alter, & Friederici, 1999; Friedrich, Kotz, Friederici, & Alter, 2004). Furthermore, a recently published study has reported an increased N400 for sentences comprising a syllable lengthening in the last word (Magne et al., 2007). However, to our knowledge the

perception of speech rhythm as a composition of the mentioned suprasegmental characteristics and expanding over a longer period of time such as the duration of a sentence has never been the subject of investigation in functional brain research.

From the point of view of neuropsychological research, speech rhythm is of particular interest related to the question of functional hemispheric lateralization. Even though the preponderance of the left hemisphere for speech processing is one of the best-evidenced facts in cognitive neuroscience, there is no doubt that the right hemisphere also contributes to speech perception (Jung-Beeman, 2005). However, the exact function of the right hemisphere in speech processing has yet to be elucidated. The auditory association cortex, which is situated in the posterior supratemporal plane, has mostly been associated with a variety of auditory functions (Warren, Wise, & Warren, 2005), amongst others with speech melody processing (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Gandour et al., 2004). Thus, one might hypothesize that speech rhythm processing, being classified as a prosodic aspect of speech might also be related to a right temporal lobe function, more specifically the posterior superior temporal gyrus (pSTG). Support for this reasoning is provided by a current model of functional lateralization in language processing – the "asymmetric sampling in time" (AST) hypothesis (Poeppel, 2003). This model assumes that auditory fields in the two hemispheres prefer different temporal integration windows. Basically, the AST suggests that perception of rapidly changing cues (~40 Hz) preferentially drives the left hemisphere while the right auditory cortex is better adept at processing slowly changing acoustic cues (~4 Hz) available in spoken language. It is further proposed that the posterior portion of auditory association cortex is the candidate region that accommodates this temporal processing. Speech rhythm involves the integration of rhythmic speech elements throughout a longer period of time such as the duration of a sentence. Thus we hypothesize that speech rhythm perception should be considered a suprasegmental process that is preferentially supported by brain structures of the right hemisphere.

While investigating lateralization effects in language processing a second line of

research needs to be considered. It has been demonstrated by auditory functional imaging studies that the pattern of hemodynamic responses in frontal (Hsieh, Gandour, Wong, & Hutchins, 2001) but most interestingly also in temporal cortical areas (Brechmann & Scheich, 2005; Plante, Creusere, & Sabin, 2002; Tervaniemi & Hugdahl, 2003) depends not only on the particular stimulus class but also on the task performed. Interestingly, this task-dependent modulation usually refers to a lateralization shift in auditory cortical areas involved in speech processing (Noesselt, Shah, & Jancke, 2003; Scheich, Brechmann, Brosch, Budinger, & OHl, 2007; Tervaniemi & Hugdahl, 2003). Authors reporting task-dependent modulation effects have associated this effect with top-down processing, an interpretation that is further sustained by anatomical and functional plastic changes observed due to environmental and training influences (Tervaniemi & Hugdahl, 2003; Luders, Gaser, Jancke, & Schlaug, 2004; Josse, Mazoyer, Crivello, & Tzourio-Mazoyer, 2003). Assuming that lateralization in the temporal cortex might be influenced by task effects we consider it very important to assess speech rhythm processing under different processing conditions.

The goal of the present study was to identify specific brain structures involved in the performance of a speech rhythm task. Furthermore, the study aimed to address the question of brain activation involved in speech rhythm processing in two different processing conditions, namely explicit and implicit processing. Based on the aforementied studies we hypothesize involvement of the right perisylvian cortex in explicit speech rhythm processing.

**Material and methods**

We investigated the processing of speech rhythm by using natural pseudo-sentences spoken with different rhythmic patterns. Haemodynamic responses were measured during explicit and implicit auditory processing.

*Participants*

Twenty-five subjects participated in this study. Twenty-four subjects, equally balanced over experimental groups, were analyzed (8 woman, mean age 27.4, SD=

9.3). One subject had to be excluded from the analysis due to a performance rate of below 63.8 percent correct answers. All participants were right-handed according to the Annett-Handedness-Questionnaire (Annett, 1992). They were native speakers of Swiss German with no history of neurological, major medical, psychiatric or hearing disorders. Furthermore, subjects gave written consent in accordance with procedures approved by the local ethics committee and were paid for participation.

*Stimuli*

The stimulus material comprised a total of 144 German pseudo-sentences with natural speech rhythm manipulations. Pseudo-sentences follow the correct German syntax and phonology containing phontactically legal pseudo-words instead of content words (Friederici, Meyer, & von Cramon, 2000). As mentioned in the introduction rhythmic distinctions can be found between languages. German is considered a stress-timed language that means it is the characteristic temporal distribution of accented syllables which constitutes its speech rhythm. This temporal distribution of accented syllables, where accents are marked by modulation of pitch or intensity, is not exactly isochronous in the spoken language rather it is modulated by syllable lengthening and pauses. Notwithstanding, the German language shows a strong tendency towards isochrony (Völtz, 1991; Noel Aziz Hanna, 2003). This is especially evident in verse where speech becomes even more isochronous. We can conclude that spoken German can be described on a continuum between isochrony and non-isochrony, whereas verse is often isochronous. For the purpose of this study, we decided to use sentences with a normal conversational speech rhythm, later referred to as "nonisochronously" spoken sentences, as well as "isochronously" spoken sentences."

Stimuli were controlled for syntactic differences across experimental conditions. The latter is supposed to represent an exaggerated form of German speech rhythm. Additionally, the "isochronously" spoken sentences followed a regular meter (i.e. iambs, trochees, dactyls) and “non-isochronously” spoken sentences followed an irregular meter (i.e. iambs or trochees with a dactyl interposed between two metrical feet).

Examples: “Der Speiter pongt den spiten Galtung” = „isochronous“

x ¦X x¦X x ¦ X x¦ X x¦

„ Der Jüfele knelt den furten Pflaster“ =„non-isochronous“

x ¦X x x¦X x ¦ X x¦ X x¦

The trained female speaker was instructed to say the “non-isochronous” sentences with a normal conversational speech rhythm. Before recording the “isochronous” sentences, the speaker listened to a metronome beat of approximately the average tempo of the “non-isochronous” sentences. To ensure that the “isochronous” condition stayed an ecological condition, the speaker did not have a constant reference to the metronome while speaking. Instead, she was instructed to follow the original beat as precisely as possible without making the utterance unnatural. Additionally, pseudosentences were recorded with the intonation contour of either a question or a statement balanced over the two rhythmic conditions (Tab 1).

Additionally, a low-level auditory control condition consisting of isochronous syllables (e.g. “da de di do du”; n=36) as well as a total of 40 null events as baseline condition were randomly included in the time course of the experiment.

All stimulus items were normalized in amplitude to 70 percent of the loudest signal in a stimulus item. Then, all stimulus items were analyzed by means of the PRAAT speech editor (Boersma & Weenink, 2000) for mean duration, mean intensity (dB) and intensity on a root-means-square based measure. Stimuli were balanced with respect to mean duration and intensity on the root-means-square based measure over experimental condition. The isochronous stimuli showed a mean intensity of 68.7 dB, whereas the “non-isochronous” sentences showed a mean intensity of a 67.4 dB.

*Experimental Groups / Task*

Subjects were randomly assigned to two different experimental groups each group having to perform a different categorization task on the same set of stimuli (Tab1). The “explicit processing” group (n=12) had to judge, whether they had heard an "isochronous" or "non-isochronous" pseudo-sentence (rhythm task). The “implicit processing” group (n=12) had to judge whether they had heard a statement or question (prosody task). Subjects indicated their response by pressing a button with

the index and middle finger respectively of their right hand. Subjects of the implicit processing group were naïve with respect to the rhythmic manipulation and did not detect a rhythmic difference between the stimuli throughout the experiment as confirmed by a short debriefing after the experiment. No feedback was given during the experiment.

| Processing modus / sentence type | explicit ( n=12) | Implicit (n=12) |
|---|---|---|
| Isochronous (n = 72) | question / statement | question / statement |
| non-isochronous (n = 72) | question / statement | question / statement |

**Tab 1:** Experimental conditions

*Procedure*

In a short training session conducted prior to the fMRI experiment, subjects were made familiar with the task. We used an MR-compatible piezoelectric auditory stimulation system incorporated into standard Philips headphones for binaural stimulus delivery. Stimuli were presented in pseudorandom order binaurally. A total of 220 trials (144 pseudosentences, 36 low-level auditory control items, 40 null events) were presented in two runs. Before each stimulus presentation a fixation cross was presented for 500 ms.

*fMRI Design*

We implemented a clustered sparse temporal acquisition technique that combines the principles of a sparse temporal acquisition with a clustered acquisition of three consecutive volume scans per trial (Schmidt et al., 2007; Zaehle et al., 2007). We collected three consecutive volumes in order to cover the peak of the event-related hemodynamic signal.

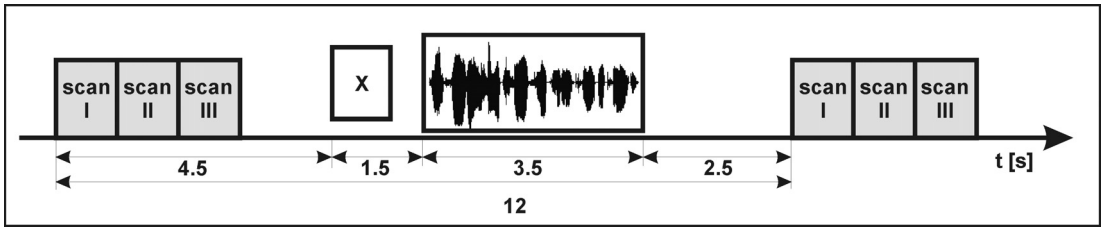



**Fig 1:** Acquisition Scheme: the figure demonstrates the three time points of acquisition (TP1, TPII, TPIII) and the stimulus presentation in a single trial.

*Data acquisition*

Data were collected using a Philips Intera 3T whole body MR unit (Philips Medical System Best, The Netherlands) equipped with an eight-channeled Philips SENSE head coil. Functional time series were collected from 16 transverse slices covering the entire peri-sylvian cortex with a spatial resolution of 2.7 x 2.7 x 4 mm using a Sensitivity Encoded (SENSE) (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999) single-shot gradient-echo planar sequence (acquisition matrix 80 x 80 voxels, SENSE acceleration factor R = 2, FOV = 220 mm, TE = 35 ms and inter-slice gap 2 mm). Three volumes were acquired per trial with each a Tacq=1000 ms, $\theta = 68^{o}$ (decay sampling) and 12 s inter cluster interval (ICI). Furthermore, we collected a standard 3D T1-weighted scan for anatomical reference with 1 x 1 x 1.5 mm spatial resolution (180 axial slices, acquisition matrix 224 x 224 voxels, TE = 2.3 ms, TR = 20 ms, $\theta = 20^{0}$).

**Data analysis**

*Behavioural data*

During the experiment the behavioral performance of the two experimental groups was measured. Data were corrected for outliers (>2 std above or below mean value). Behavioral measures were aggregated by participants and conditions. As a measure for accuracy of discrimination, the mean percentage of correct answers over all experimental conditions (without low-level auditory control condition) was calculated. An independent sample t-test was performed to identify group differences.

*fMRI data*

To account for different T1 saturation effects in subsequent volumes, we subjected the three volume scans (TP I-III) collected during each cluster to three separate analyses during pre-processing and individual statistics. The functional imaging data processing was carried out using MATLAB 6.5 (Mathworks Inc., Natiek, MA, USA) and the software package SPM99 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Functional data were realigned to the first volume, corrected for motion artifacts and normalized into standard stereotactic space (voxel size 2 x 2 x 2 mm, template provided by the Montreal Neurological Insitute (Evans et al., 1992). For spatial smoothing we applied an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Low-frequency drifts were removed using a temporal high-pass filter (mean cut-off of 307 seconds). Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations, performed separately on each voxel (Friston et al., 1995). Single trials were treated as epochs and modelled by means of a box car function. Conditions were compared by calculating contrasts between conditions of interest for each participant and time point of acquisition. Direct contrasts between isochronous and non-isochronous sentences as well as between experimental conditions and the low-level auditory control condition were calculated. We limited the presentation of results to TPII as this acquisition time is supposed to reflect the amplitude peak of the hemodynamic response as it has been done formerly (Meyer, Baumann, Marchina, & Jaencke, 2007; Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005). Contrast images were submitted to a second level group analysis. Random-effects analysis consisted of between and within group comparison. The between-group analysis consisted of a two-sampled t-test using “all sentences>auditory control” contrasts. The within-group analysis consisted of a paired-t-test using “isochronous sentences>auditory control” and “non-isochronous sentences>auditory control” contrasts. Results were thresholded at $T = 3.50$ for the two-sampled t-test and $T = 4.02$ for the paired t-tests ($p = 0.001$, uncorrected for multiple comparisons). Only clusters of significant size ($p< 0.05$, corrected for multiple comparison with a spatial extent of $k = 45$ for the two-sampled t-test and $k = 27$ for the paired t-test) were reported (Worsley et al., 1996).

**Results**

*Behavioral results*

Due to technical problems, the performance data of 4 subjects, two of each group could not be recorded. The implicit processing group (prosody task, n=10) performed better than the explicit processing group (rhythm task, n = 10). 80.25% of the answers were correct for the explicit processing group and 98.6% for the implicit processing group. A two-sampled t-test performed to identify group differences revealed a significant difference in response accuracy ($t_{(18)}$=-5.28; p< 0.000).

*FMRI results*

The present paper addressed two main issues. First we were interested in the neurofunctional correlate related to the performance of the speech rhythm task. Second we aimed to analyze the brain areas involved in the processing of exaggerated German speech rhythm in an explicit and an implicit processing condition.

Performance of the speech rhythm task

Between-group comparison between explicit and implicit processing of all sentences (Fig 2, Tab 2) revealed activation the SMA for the explicit processing group compared to the implicit processing group. This activation was found bilaterally, although the peak activation was in the right hemisphere. The cluster extended into the medial part of the superior frontal gyrus and the right cingulate gyrus. Furthermore, this contrast revealed bilateral activation in the left and right insula encroaching onto the neostriatum, namely the Putamen in the right hemisphere. In the right hemisphere we found an additional activation in the inferior frontal gyrus (pars opercularis).

Explicit processing of isochronous speech rhythm

To reveal the effect of speech rhythm we compared isochronously spoken sentences to non-isochronously spoken sentences in the explicit processing condition (Fig 3, Tab 3) in a direct contrast. We observed statistically significant activation in right

superior temporal gyrus extending into the supramarginal gyrus and the Parietal Operculum.

Implicit processing of isochronous speech rhythm

The contrast between isochronously spoken and non-isochronously spoken sentences in the implicit processing condition (Fig 3, Tab 4) revealed activation in the left hemisphere, namely the supramarginal gyrus and the superior temporal gyrus extending into the parietal Operculum for the isochronously spoken sentences compared to non-isochronically spoken sentences.

Taken together, direct contrast between isochronous and non-isochronous sentences revealed a substantial effect observed for the processing of isochronously spoken speech, which was differentially lateralized in the explicit as compared to the implicit processing condition. Furthermore, the performance of the speech rhythm task relied on activation of the supplementary motor area (SMA) as well as the insula (INS) bilaterally and the right inferior frontal gyrus (IFG).

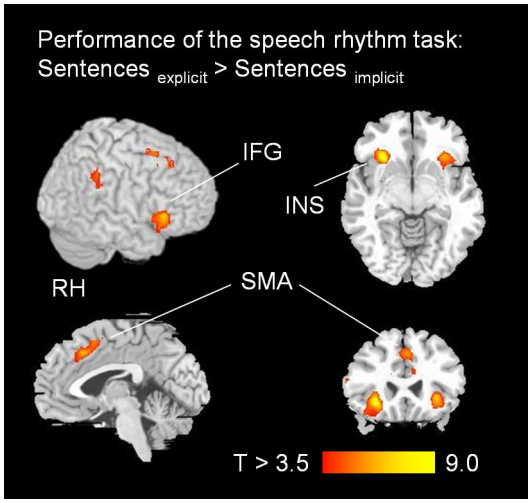



**Fig 2:** Brain areas that showed significantly greater activation during the performance of the rhythm task (explicit) as compared to the implicit processing condition. The figure shows resulting T-maps of the two-sampled t-test using sentences vs. auditory control contrast for the explicit as compared to the implicit processing condition (up right z= -7, down left x= 2, down right y = 25).The top-left image shows the activations projected on the cortical surface. Activation is thresholded at T = 3.50 ($p<0.001$, uncorrected) and cluster corrected for multiple comparisons at $p<0.05$ (Worsley et al. 1996). Figures are displayed in neurological convention.

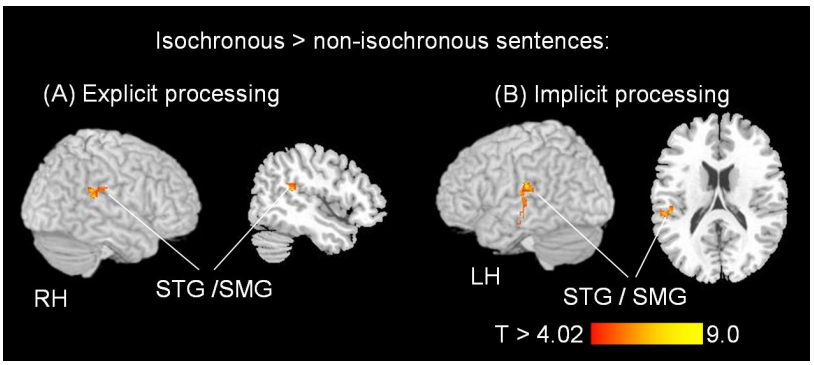



**Fig 3:** Brain areas showing significantly greater activation during the processing of isochronous compared to non-isochronous sentences in (A) the explicit processing condition (right x = 48) and (B) the implicit processing condition (right z = 18). For both the explicit and implicit processing condition the activation in the sagital view is projected on the lateral convexity of cortical surface. Each cluster is thresholded at T = 4.02 ($p<0.001$, uncorrected) and cluster corrected for multiple comparisons at $p<0.05$ (Worsley et al. 1996). Figures are displayed in neurological convention.

**Tab 2:** Performance of the speech rhythm task

| | Left hemisphere | | | | Right hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| Anatomical description | p value* | T value | cluster size | x,y,z [mm] | p value* | T value | cluster size | x,y,z [mm] |
| INS extending into NS | | | | | 0.000 | 7.57 | 212 | 27 21 -9 |
| IFG (pars opercularis ) | | | | | 0.000 | 6.89 | 168 | 51 12 12 |
| SMG | | | | | 0.000 | 6.38 | 257 | 42 -42 36 |
| INS extending into IFG (pars orbitalis/triangularis) | 0.000 | 6.3 | 176 | -30 24 -3 | | | | |
| SMA and cingulate gyrus extending into SFG | 0.000 | 6.12 | 249 | 0 21 42 | | | | |
| IPL | 0.041 | 4.89 | 46 | -39 -42 42 | | | | |

Two-sampled t-test: (all sentences> auditory control) $_{expl}$ > (all sentences >auditory control ) $_{impl}$; Coordinates are given according to the MNI-space; * p-values corrected for entire volume on cluster-level (k=45)

*Legend:* INS = Insula, NS = neostriatum (caudate/head, putamen), IPL = inferior parietal lobe, Ppar= planum parietale, RO= rolandic operculum, STS = superior temporal gyrus, SFG = superior frontal gyrus, OG = occipital gyrus, PP = planum polare, PT = planum temporale, SmA = supplementary motor area, IPG= inferior parietal gyrus, SMG = supplementary motor area, PaOp = parietal Operculum, IFG = inferior frontal gyrus

**Tab 3:** Explicit isochronous speech rhythm processing

| | Left hemisphere | | | | Right hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| Anatomical description | p value* | T value | cluster size | x,y,z [mm] | p value* | T value | cluster size | x,y,z [mm] |
| STG extending into SMG / PaOp | | | | | 0.002 | 6.72 | 49 | 48 -39 18 |

Paired t-test: (Isochronous sentences > auditory control) > (Non-isochornous sentences > auditory control); Coordinates are given according to the MNI-space; * p-values corrected for entire volume on clustser level (k = 27).

**Tab 4:** Implicit isochronous speech rhythm processing

| | Left hemisphere | | | | Right hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| Anatomical description | p value* | T value | cluster size | x,y,z [mm] | p value* | T value | cluster size | x,y,z[mm] |
| SMG extending into STG / PaOp | 0.000 | 7.45 | 75 | -57 -30 21 | | | | |

Paired t-test: (isochronous sentences > auditory control) > (Non-isochronous sentences > auditory control); Coordinates are given according to the MNI-space; * p-values corrected for entire volume on clustser level (k = 27).

## Discussion

The main goal of the study was to identify brain areas related to speech rhythm processing, that is, the performance of a speech rhythm task. A further objective was to determine possible differences between explicit and implicit rhythm perception. In the following section we will first discuss the brain activation evoked by the performance of the speech rhythm task. Then we will discuss the differential effects found for isochronous processing as compared to non-isochronous processing in both the explicit and subsequently in the implicit processing condition.

*Performance of the speech rhythm task*

A clear effect related to the explicit performance of the rhythm task was observed in the *mesial premotor cortex (SMA-pre/proper)*, in contrast to the implicit processing group. We suggest that this finding gives rise to the notion of a general function of the mesial premotor cortex which is related to timing processes in auditory perception as described in the following section. Considering other recent findings this result speaks for a multimodal function of the SMA in temporal perception.

The SMA has frequently been reported to be involved in motor processes such as

finger movement or speech production (Riecker et al., 2005). In the context of motor research the SMA is divided into pre-SMA, associated with motor preparation, and SMA-proper, associated with movement execution, encoding and retrieval of motor sequences (Cunnington, Windischberger, Robinson, & Moser, 2005; Lutz, Specht, Shah, & Jancke, 2000; Lau, 2006; Lewis, Wing, Pope, Praamstra, & Miall, 2004; Koeneke, Lutz, Wustenberg, & Jancke, 2004; Kansaku et al., 2005). However, there is evidence that this involvement of the SMA-proper as well as the pre-SMA in motor processes might be due to a specific function in the timing of motor actions, a function which might not only be restricted to motor processing. For example, both areas have been found to show an increasing relation between complexity, that is temporal complexity, of motor sequences and activation (Lewis et al., 2004; Lutz, Specht, Shah, & Jancke, 2000; Bengtsson, Ehrsson, Forssberg, & Ullen, 2005; Riecker, Kassubek, Gröschel, Grodd, & Ackermann, 2006).

The assumption of a temporal processing function of the SMA is sustained by evidence from perception research in the visual domain, which described a correlation between attention to time and activity in the SMA (Coull, 2004a; Nobre & O'Reilly, 2004; Lewis & Miall, 2003). A similar finding has been reported in the sensorimotor domain showing an involvement of the SMA in time estimation after a specific stimulation (Macar, Anton, Bonnet, & Vidal, 2004b). Interpreting the above mentioned studies in the direction of a multimodal involvement of the SMA in time perception (Schubotz, Friederici, & von Cramon, 2000; Pastor, Macaluso, Day, & Frackowiak, 2006), our study poses an enhancement of this assumption in that we used acoustic stimuli and found a very specific activation in the SMA related to temporal perception. Further research will specify the different processes underlying timing perception. Furthermore, our results found in the contrast between sentences and low-level auditory control condition, clearly speak for a task-driven effect of sequential timing perception.

One might argue that the two tasks are not of equal difficulty as suggested by the performance rate and that the SMA activity might be induced by this difference. A recently published study argues that SMA in all timing tasks is not related to timing but rather to supplementary activities (Livesey, Wall, & Smith, 2007). On the other hand, studies that aimed to disentangle task difficulty and temporal processing

associated activity in the pre-SMA and SMA proper specifically with temporal processing (Macar, Anton, Bonnet, & Vidal, 2004a; Coull, 2004b). Moreover, studies investigating task difficulty in a pitch discrimination task did not report an effect in SMA (Reiterer et al., 2005). In our opinion, this is an indication that task difficulty is not significantly involved in our data.

An alternative interpretation of the SMA activation could be given in terms of strategy mechanisms. It is possible that the SMA activation relates to spontaneous and intuitive synchronization of rhythm perception and body movements. That is, subjects intuitively "feel" the rhythm in their body while listening to the specific sentences. This assumption is sustained by reports showing that auditory cues are very effective in facilitating accurate rhythmic body movements (Patel, Iversen, Chen, & Repp, 2005) as well as by comments made by subjects after the performance of the experiment. This sensation could be mediated by motor cortices such as the SMA. However, spontaneous motor synchronization mechanisms must be a topic of further investigations.

Furthermore, our results show an extensive activation in the *insula of both hemispheres* and in the *right inferior frontal gyrus*, specifically the *right pars opercularis*. Besides its function in motor (Jantzen, Steinberg, & Kelso, 2005) and speech production processes (Ackermann & Riecker, 2004; Bohland & Guenther, 2006; Blank, Scott, Murphy, Warburton, & Wise, 2002), specifically in motor timing, the insula has repeatedly been found in various auditory processes (Bamiou, Musiek, & Luxon, 2003; Giraud et al., 2004; Mutschler et al., 2007). Of note are auditory temporal processing (Ackermann et al., 2001; Lewis, Beauchamp, & DeYoe, 2000; Lewis et al., 2003), speech melody perception (Meyer et al. 2002), and auditory feedback and pitch control (Zarate & Zatorre, 2005) which have recently been reported. Moreover, functional lateralization in the insulae has been suggested on the basis of temporal acoustic characteristics (Ackermann et al., 2001) as well as on the function of pitch-related stimuli (Wong, Parsons, Martinez, & Diehl, 2004). Very recently Bamiou et al. (Bamiou et al., 2006) gave strong support to the notion of a crucial role of the insula in temporal processing, in particular, temporal resolution and sequencing. Thus, there is an evolving view that the insula is crucial

in auditory timing perception. Our processing task involves the perception of the specific accent patterns, that is, the sequencing of syllables. Therefore, our finding of an insular involvement in the sequencing task is not surprising and in line with the above mentioned hypothesis about insular functions.

Of particular note is the activation in the right insula extending into the *basal ganglia (Putamen)*. Lesion studies as well as imaging studies revealed an involvement of the basal ganglia in the processing of suprasegmental speech cues, namely emotional speech melody in speech production (Pell, Cheang, & Leonard, 2006) and speech melody perception (Kotz et al., 2003; Zaehle, Wustenberg, Meyer, & Jancke, 2004). Furthermore, lesion studies have found that patients with a basal ganglia insult also show symptoms of impaired prosodic functions (Van Lancker Sidthis D., Pachana, Cummings, & Sidthis, 2006). A very recent finding has indicated a role of basal ganglia in auditory beat perception (Grahn & Brett, 2007). These data expand the function of the basal ganglia over the commonly reported functions in motor processing. Moreover our data strongly suggest that the involvement of basal ganglia in the processing of auditory suprasegmental cues might not be limited to melodic aspects of speech but be necessary for the perception of rhythmic aspects of speech as well.

We want to put an additional emphasis on the activity observed in the right inferior frontal gyrus. The right IFG has repeatedly been reported in addition to left frontal brain areas in speech processing and associated with increasing sentence complexity (Michael, Keller, Carpenter, & Just, 2001), syntax processing (Ni et al., 2000; Meyer, Friederici, & von Cramon, 2000; Poldrack et al., 2001), or the performance of specific speech tasks (Chou et al., 2006; Meyer, Friederici, & von Cramon, 2000). However, there is also a considerable amount of research which has associated the right IFG with specific acoustic perception, i.e. pitch processing. Activation in the right IFG has been found during pitch processing in both non-linguistic (Liu et al., 2006; Humphries, Willard, Buchsbaum, & Hickok, 2001; Hsieh, Gandour, Wong, & Hutchins, 2001) and speech contexts (Gandour et al., 2004; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Pitch processing plays a crucial role in speech processing of tonal languages, that is languages where the pitch of a word carries a specific semantic content

(Gandour et al., 2003). In non-tone languages pitch becomes important in the context of speech melody perception (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002). In German language the perception of accent patterns includes both, pitch and intensity patterns. Thus, in our study the activation in the right IFG must be interpreted more generally in relation to the integration of accent patterns. Furthermore, in our specific task, subjects processed the stimuli focussing on the accent pattern independently of a linguistic context. Thus, we strongly believe that our data provide evidence, that the right IFG can be selectively activated in tasks requiring higher suprasegmental cue processing.

*Explicit and implicit processing of isochronous speech rhythm*

We analyzed the effect of speech rhythm, isochronous vs. non-isochronous sentences, in both the explicit and the implicit processing condition. This analysis is based on recent findings of differentially lateralized effects in auditory processing depending on the context of the stimulus perception (Brechmann et al., 2005) as well as on findings reporting a top-down modulation on auditory processing (Tervaniemi & Hugdahl, 2003).

In the explicit processing condition we measure task-driven processing of isochronous speech. Interestingly, this process leads to temporal activation in the right hemisphere, namely activation in the *posterior part of the right STG* (specifically planum temporale, PT) extending into the supramarginal gyrus, for the isochronous as compared to the non-isochronous sentences. Most theories of the functional role of the PT now assume that the PT in the left and the right hemisphere is involved in a parameter specific segregation of auditory sounds leading to differentially lateralized processes of the two hemispheres. Two different parameter-specific concepts of the PT are found in the literature, however they are widely considered to be two sides of the same coin. One concept posits that the posterior part of the STG is sensitive to spectrotemporal auditory processing and associates the right STG with spectral and the left STG with temporal auditory processing (Zatorre & Belin, 2001; Menon et al., 2002; Meyer et al., 2005; Griffiths & Warren, 2002; Zaehle et al. 2004). A complimentary model, the AST, specifies the potential functions of the pSTG slightly differently (Poeppel, 2003). This model proposes that

the left non - primary auditory cortex is preferentially driven by rapidly changing acoustic cues, while right auditory cortex is more amenable to slowly changing acoustic cues. The perception of suprasegmentally modulated speech rhythm that unfolds in the time range of syllables (150-200 ms) should, therefore, more likely be supported by the right posterior temporal plane. Thus we state that our findings can be explained by the AST of functional lateralization. However, our results have a second highly important implication. We find the right hemisphere to be involved in the processing of isochronous speech rhythm in the explicit processing condition only. The same contrast in the implicit processing condition does not show right hemispheric activation. Therefore, our results additionally provide strong evidence for attention modulated activation in the auditory cortex. More precisely, we believe that this gives strong evidence for a top-down modulation on lateralization effects in the auditory cortex (Brechmann et al., 2005; Tervaniemi & Hugdahl, 2003).
One might reason why we did not find activation in the anterior superior temporal cortex of the right hemisphere in response to speech rhythm perception. Such activation has been found in melody perception and production (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Brown, Martinez, Hodges, Fox, & Parsons, 2004), as well as in the perception of non-speech vocal sounds (Belin, Zatorre, & Ahad, 2002). We argue that speech rhythm relies strongly on linguistic information, which has not been found to involve the right anterior superior temporal brain areas.

In the implicit processing condition we measure unattended processing of speech rhythm. Furthermore, as confirmed after the completion of the experiment subjects were not aware of the rhythmic manipulation within the sentences. Surprisingly, the contrast between isochronous and non-isochronous sentences revealed significant activation in the *left supramarginal gyrus* extending into the *left posterior superior temporal gyrus (STG)*, more specifically the planum temporale (PT) and extending in the left parietal Operculum for the isochronous condition as compared to the non-isochronous condition. The extent of activation observed in the left posterior STG encroached onto the upper-posterior part of the Sylvian fissure and the supramarginal gyrus. The extension of activity over these cortical areas is not surprising. It has been

shown that subjects with a larger left PT tend to show a larger leftward functional asymmetry of several perisylvian areas, namely the inferior parietal lobule outside the supramarginal gyrus, Heschl's gyrus, the Rolandic operculum, and the temporal pole (Josse, Mazoyer, Crivello, & Tzourio-Mazoyer, 2003). This finding suggests possible functional common ground between these cortex areas. Hickok and collegues (Hickok, Buchsbaum, Humphries, & Muftuler, 2003) first described activation in the parietal operculum and in the adjacent posterior part of the Sylvian fissure, at the parietal-temporal boundary for both speech listening tasks and covert rehearsal of spoken language. The authors suggested that this area might be involved in auditory-motor activity similar to the visual-motor response properties in the dorsal visual stream (Rizzolatti, Fogassi, & Gallese, 1997). In the context of our study it is important to take notice of another line of research reporting that musical meter, as the whole-numbered (i.e. isochronous) sequence of accented acoustic events, is the characteristic feature of music which relates to body movements (Lerdahl & Jackendoff, 1983). Very recently Chen et al. (Chen, Zatorre, & Penhune, 2006) investigated the neural correlate of a synchronization process of body movements using the meter of an auditory cue. The authors found the left planum temporale, the right posterior STG and bilateral dorsal premotor cortex to covariate with metric salience and thus to facilitate the synchronization of motor actions to auditory cues. Therefore, we suggest that listening to exaggerated speech rhythm in the implicit processing condition leads to involuntary auditory-motor interactions due to the salience of isochronously spoken sentences.

**Conclusion**

In the context of an fMRI-study on speech rhythm perception we found the SMA, right IFG, bilateral insulae as well as the right basal ganglia to be involved in sequential timing perception related to the performance of a rhythmic categorization task on speech stimuli. These data provide evidence for an involvement of the SMA in the perception of time intervals across modalities and for a specific task-related function of the right inferior frontal gyrus in the processing of accent patterns in speech.

Furthermore, we observed a clear effect of speech rhythm, which was lateralized to

the right auditory cortex in an explicit, task-driven processing condition. This observation can be explained by the fact that suprasegmental speech rhythm perception occurs over a long temporal integration window. Interestingly, this right lateralized activation did not occur in an implicit processing condition. This finding indicates that right auditory association cortex activity can be modulated by top-down processing and can be selectively recruited depending on task demands. The same contrast in an implicit, stimulus-driven processing condition evoked differentially lateralized brain activation in the left hemisphere. This finding is assumed to reflect an involuntary auditory-motor interaction effect.

Acknowledgments

This work was supported by the Swiss National Foundation (SNF-grant 3200B0-105877 conferred to M. Meyer). We are grateful to the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig for assistance in stimulus recording. We specially thank Dr. Dorothea Weniger from the University Hospital of Zurich and Elsa Sägesser for helpful comments on previous versions of this manuscript. Furthermore, we thank Dr. Conny F. Schmidt from the Institute for Biomedical Technology, ETH and University of Zurich, Switzerland, for assistance in the data acquisition.

Reference List

Ackermann, H. & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang, 89,* 320-328.

Ackermann, H., Riecker, A., Mathiak, K., Erb, M., Grodd, W., & Wildgruber, D. (2001) Rate-dependent activation of a prefrontal-insular-cerebellar network during passive listening to trains of click stimuli: an fMRI study. *Neuroreport, 12,* 4087-4092.

Annett, M. (1992). 5 Tests of Hand Skill. *Cortex, 28,* 583-600.

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Res.Brain Res.Rev., 42,* 143-154.

Bamiou, D. E., Musiek, F. E., Stow, I., Stevens, J., Cipolotti, L., Brown, M. M. et al. (2006). Auditory temporal processing deficits in patients with insular stroke. *Neurology, 67,* 614-619.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res.Cogn Brain Res., 13,* 17-26.

Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., & Ullen, F. (2005). Effector-independent voluntary timing: behavioural and neuroimaging evidence. *Eur.J.Neurosci., 22,* 3255-3265.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125,* 1829-1838.

Boersma, P. & Weenink, D. (2000). PRAAT: Doing phonetics by computer. [Computer software]. Instiute of phonetic Sciences, University of Amsterdam.

Bohland, J. W. & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*.

Bolinger, D. (1965). *Pitch accent and sentence rhythm, forms of English: accent, morpheme, order.* Cambridge MA: Harvard University Press.

Brechmann, A. & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex, 15,* 578-587.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Brain Res.Cogn Brain Res., 20,* 363-375.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage., 26,* 1119-1127.

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*.

Chou, T. L., Booth, J. R., Burman, D. D., Bitan, T., Bigio, J. D., Lu, D. et al. (2006). Developmental changes in the neural correlates of semantic processing. *Neuroimage., 29,* 1141-1149.

Coull, J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. *Brain Res.Cogn Brain Res., 21,* 216-226.

Cunnington, R., Windischberger, C., Robinson, S., & Moser, E. (2005). The selection of intended actions and the observation of others' actions: A time-resolved fMRI study. *Neuroimage*.

Curtin, S. M. T. H. C. M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96,* 233-262.

Cutler, E. A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. In Altmann G.T.M. (Ed.), *Psycholinguistics: Critical Concepts.* London: Routledge. (pp. 157-177).

Evans, A. C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W. et al. (1992). Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage., 1,* 43-53.

Frazier, L., Carlson, K., & Clifton Jr., C. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn Sci., 10,* 244-249.

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang, 75,* 289-300.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Brain Res.Cogn Brain Res., 20,* 300-308.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *J.Cogn Neurosci., 16,* 541-552.

Friston, K. J., Holmes, A. P., Worsley, K. P., Poline, J. B., Frith, C. D., & Frackowiak, R. S. (1995). Statistical parameter maps in functional imaging: A general linear approach. *Hum.Brain Mapp., 2,* 189-210.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y. et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage., 23,* 344-357.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X. et al. (2003). Neural correlates of segmental and tonal information in speech perception. *Hum.Brain Mapp., 20,* 185-200.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C. et al. (2004). Contributions of Sensory Input, Auditory Search and Verbal Comprehension to Cortical Activity during Speech Processing. *Cereb.Cortex, 14,* 247-255.

Grahn, J. A. & Brett, M. (2007). Rhythm and Beat Perception in Motor AReas of the Brain. *Journal of Cognitive Neuroscience, 19,* 893-906.

Griffiths, T. D. & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci., 25,* 348-353.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J.Cogn Neurosci., 15,* 673-682.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci.Res., 54,* 276-280.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001a). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang, 76,* 227-252.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001b). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang, 76,* 227-252.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport, 12,* 1749-1752.

Jantzen, K. J., Steinberg, F. L., & Kelso, J. A. (2005). Functional MRI reveals the existence of modality and coordination-dependent timing networks. *Neuroimage., 25,* 1031-1042.

Josse, G., Mazoyer, B., Crivello, F., & Tzourio-Mazoyer, N. (2003). Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. *Brain Res.Cogn Brain Res., 18,* 1-14.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends Cogn Sci., 9,* 512-518.

Kansaku, K., Muraki, S., Umeyama, S., Nishimori, Y., Kochiyama, T., Yamane, S. et al. (2005). Cortical activity in multiple motor areas during sequential finger movements: an application of independent component analysis. *Neuroimage, 28,* 669-681.

Koeneke, S., Lutz, K., Wustenberg, T., & Jancke, L. (2004). Bimanual versus unimanual coordination: what makes the difference? *Neuroimage, 22,* 1336-1350.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang, 86,* 366-376.

Lau, H. R. R. D. P. R. E. (2006). Dissociatin response selection and conflict i the medial frontal surface. *Neuroimage, 29,* 446-451.

Lee, C. S. & Todd, N. P. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition, 93,* 225-254.

Lehiste, I. (1977). Isochrony Reconsidered. *Journal of Phonetics, 5,* 253-263.

Lerdahl, F. & Jackendoff, R. (1983). An Overview of Hierarchical Structure in Music. *Music Perception, 1,* 229-252.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison ofvisual and auditory motion processing in human cerebral cortex. *Cereb.Cortex, 10,* 873-888.

Lewis, P. A. & Miall, R. C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia, 41,* 1583-1592.

Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., & Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia, 42,* 1301-1312.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K. et al. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage., 29,* 515-523.

Livesey, A. C., Wall, M. B., & Smith, A. T. (2007). Time perception: manipulation of task difficulty dissociates clock functions from other cognitive demands. *Neuropsychologia, 45,* 321-331.

Luders, E., Gaser, C., Jancke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage, 22,* 656-664.

Lutz, K., Specht, K., Shah, N. J., & Jancke, L. (2000). Tapping movements according to regular and irregular visual timing signals investigated with fMRI. *Neuroreport, 11,* 1301-1306.

Macar, F., Anton, J. L., Bonnet, M., & Vidal, F. (2004a). Timing functions of the supplementary motor area: an event-related fMRI study. *Brain Res.Cogn Brain Res., 21,* 206-215.

Macar, F., Anton, J. L., Bonnet, M., & Vidal, F. (2004b). Timing functions of the supplementary motor area: an event-related fMRI study. *Brain Res.Cogn Brain Res., 21,* 206-215.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronalnd-Martinet, R., & Besson, M. (2007). Influence of Syllabic Lenghening on Semantic Processing in Spoken French: Behavioral and Electrophysiological Evidence. *Crebral Cortex Advanced Access*.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language, 45,* 103-132.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D. et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage., 17,* 1742-1754.

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum.Brain Mapp., 17,* 73-88.

Meyer, M., Baumann, S., Marchina, S., & Jaencke, L. (2007). Haemodynamic responses in human mutlisensory and auditory association cortex to purely visual stimulation. *BMC Neuroscience, 8(14). doi: 10.1186/1471-2202-8-14.*

Meyer, M., Friederici, A. D., & von Cramon, D. Y. (2000). Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. *Brain Res.Cogn Brain Res., 9,* 19-33.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang, 89,* 277-289.

Meyer, M., Zaehle, T., Gountouna, V. E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport, 16,* 1985-1989.

Michael, E. B., Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. *Hum.Brain Mapp., 13,* 239-252.

Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. (2007). A Rapid Sound-Action Association Effect in Human Insular Cortex. *PLOS ONE, 2,* e259.doi:10.1371/journal.pone.0000259.

Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41,* 233-243.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E. et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *J.Cogn Neurosci., 12,* 120-133.

Nobre, A. C. & O'Reilly, J. (2004). Time is of the essence. *Trends in Cognitive Sciences, 8,* 387-389.

Noel Aziz Hanna, P. (2003). *Sprachrhythmus in Metrik und Alltagssprache*. München: Wilhelm Fink Verlag.

Noesselt, T., Shah, N. J., & Jancke, L. (2003). Top-down and bottom-up modulation of language related areas - An fMRI Study. *BMC.Neurosci., 4,* 13.

Pastor, M. A., Macaluso, E., Day, B. L., & Frackowiak, R. S. (2006). The neural basis of temporal auditory discrimination. *Neuroimage, 30,* 512-520.

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Exp.Brain Res., 163,* 226-238.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36,* 767-776.

Pell, M. D., Cheang, H. S., & Leonard, C. L. (2006). The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. *Brain Lang, 97,* 123-134.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: activation interacts with task demands. *Neuroimage., 17,* 401-410.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41,* 245-255.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M. et al. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *J.Cogn Neurosci., 13,* 687-697.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine, 42,* 952-962.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265-292.

Reiterer, S. M., Erb, M., Droll, C. D., Anders, S., Ethofer, T., Grodd, W. et al. (2005). Impact of task difficulty on lateralization of pitch and duration discrimination. *Neuroreport, 16,* 239-242.

Riecker, A., Kassubek, J., Gröschel, K., Grodd, W., & Ackermann, H. (2006). The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signal chagnes at striatal and cerebellar structures. *Neuroimage, 29,* 46-53.

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W. et al. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology, 64,* 700-706.

Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: from sight to action. *Curr.Opin.Neurobiol., 7,* 562-567.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & OHl, F. W. (2007). The Cognitive Auditory Cortex: Task-Specificity of Stmulus REpresentations. *Hearing Research*.

Schmidt, C. F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., & Jancke, L. (2007). Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. *Human Brain Mapping,* doi:10.1002/hbm.20372.

Schubotz, R. I., Friederici, A. D., & von Cramon, D. Y. (2000). Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. *Neuroimage, 11,* 1-12.

Segui, J., Djupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification and lexical access. In G.Altmann & R. Sillcock (Eds.), *Cognitive Models of Speech Processing* (pp. 263-280). Cambridge: MIT Press.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat.Neurosci., 2,* 191-196.

Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res.Brain Res.Rev., 43,* 231-246.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. *Developmental Science, 8,* 26-35.

Van Lancker Sidthis D., Pachana, N., Cummings, J. L., & Sidthis, J. J. (2006). Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. *Brain Lang, 97,* 135-153.

Völtz, M. (1991). Das Rhythmusphänomen. *Zeitschrift für Sprachwissenschaft, **10**,* 284-296.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci., 28,* 636-643.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog.Brain Res., 156,* 249-268.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J.Neurosci., 24,* 9153-9160.

Worsley, K. P., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation . *Hum.Brain Mapp., 4,* 58-73.

Zaehle, T., Schmidt, C. F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P. et al. (2007). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *Neuroimage, 37*,1195-1204.

Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *Eur.J.Neurosci., 20,* 2447-2456.

Zarate, J. M. & Zatorre, R. J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences, 1060,* 404-408.

Zatorre, R. J. & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb.Cortex, 11,* 946-953.

## Experiment II: Top-down modulation in frontal and temporal brain areas

Authors: Geiser E., Besson, M., Jancke L., & Meyer M.

**Abstract**

The present fMRI-study investigated the processing of semantic and prosodic information in speech and aimed to identify brain areas displaying interaction effects between these two speech characteristics using an event-related sparse-fMRI design. The second aim of the study was to investigate the effect that different processing conditions have on the processing of these stimulus characteristics. Stimulus material consisted of sentence stimuli which were manipulated according to a 2x2-design with the factors "prosody" (congruous and incongruous) and "semantic" (congruous and incongruous). Subjects performed both a prosodic and a semantic task on the same stimulus material. An analysis of anatomically pre-defined ROIs revealed interactions between prosodic and semantic processing takes place in the posterior middle temporal gyrus bilaterally as well as bilaterally in the posterior inferior frontal gyrus (BA44) and the planum temporale. We also demonstrate that the experimental task significantly affects the responsiveness of the posterior middle temporal gyrus to auditory stimuli in that stronger activity is found as a function of either congruency or incongruency.

**Introduction**

Speech melody is of central relevance in verbal communication. We know that correct amplitude and frequency modulation contribute to robust speech recognition (Zeng et al., 2005). In non-melodic languages such as German, French or Spanish, speech melody is used to convey important information about the speaker's intentions and emotional state. But it is also used to encode linguistic information. In tonal languages it even encodes semantic information (Gandour et al., 2003). In non-tonal languages a decrease or increase in pitch at the end of a sentence can encode a statement or question.

The involvement of right hemispheric brain areas in speech melody processing (prosody) is well evidenced in literature. Lesion studies show impairment of right hemispheric damaged individuals in whole sentence prosody processing (Baum & Dwivedi, 2003). In healthy subjects the right superior temporal gyrus has been found to be involved in the processing of pitch-manipulated sentences and of prosodic-only speech (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Lattner, Meyer, & Friederici, 2005; Friederici & Alter, 2004). Within the superior temporal gyrus, specifically the planum temporale has been suggested to be involved in the processing of prosody (Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004). In addition to processing found in temporal brain areas, in some circumstances frontal areas, namely the right inferior frontal gyrus, were found to process speech prosody. $F_o$ modulations have been reported to specifically involve right BA44 in a passive listening task to flattened and normal speech (Hesling, Clement, Bordessoules, & Allard, 2005). Interestingly, a right hemispheric lateralization of prosodic processing was even found in infants and young children (Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Wartenburger et al., 2006).

Other studies have found bilateral involvement of prosodic processing in children (Plante, Holland, & Schmithorst, 2006). In adults, bilateral inferior frontal and temporal activity was observed in the contrasts between 'melodically rising questions' and either 'melodically falling questions' or 'falling statements' (Doherty, West, Dilley, Shattuck-Hufnagel, & Caplan, 2004). Explicit processing of emotional speech relies on bilateral fronto-opercular brain regions (Kotz et al.,

2003). Thus, it appears that there are certain circumstances in which both the left hemisphere and the right hemisphere are involved in prosody processing.

One line of research assumes a differential lateralization within prosodic processing depending on the emotional content or emotional valence (Pell, 2005). On the basis of an investigation of Parkinson's patients, the author suggested that the right hemisphere is involved in emotional prosody processing while the left hemisphere is involved in the processing of prosodic aspects "as a code embedded with language content". While this finding of a specific involvement of the right superior temporal gyrus in affective prosody has been recently sustained (Beaucousin et al., 2007), a current study on emotional prosody has suggested a differential lateralization based on the expected valence of the emotion (Rodway & Schepman, 2007). Other researchers suggest differential lateralization depending on the linguistic function of prosodic speech elements. Investigations of Chinese subjects have found that while tones associated with a semantic content are processed in the left hemisphere, the processing of the same tonal stimuli by listeners naïve to Chinese language occurred in the right hemisphere (Gandour et al., 2003; Gandour et al., 2004). However, once trained to understand the lexical tone in Mandarin the latter subjects showed increased activity in the left superior temporal gyrus and in the right inferior frontal gyrus (Wang, Sereno, Jongman, & Hirsch, 2003). Thus, this evidence suggests that the involvement of cortical brain areas in prosodic processing seems to depend on emotional content and linguistic relevance. Moreover, it also depends on the task performed on the stimulus material. When active rehearsing of prosodic aspects of speech was required, the left STG became involved in addition to the right hemisphere (Meyer et al., 2004). It appears that the left hemisphere is involved task-dependently in prosody processing.

The above mentioned observations suggest the notion that the perception of prosody could depend on the task performed on specific stimuli. Task-dependent modulation of stimulus-related cortical processing has been reported earlier (Hsieh, Gandour, Wong, & Hutchins, 2001; Brechmann & Scheich, 2005; Tervaniemi & Hugdahl, 2003; Poeppel et al., 1996). However, to our knowledge this effect has not been specifically investigated for prosodic processing. Thus, one focus of our

research is the processing of prosody in different processing conditions. Several recently published studies demonstrate that sensory cortex areas (Geiser, Zaehle, Jancke, & Meyer, 2008) as well as frontal brain areas (Scheich, Brechmann, Brosch, Budinger, & Ohl, 2007) display a different functional lateralization depending on the context of perception. Thus, more specifically, we assume a differential effect of processing mode on the hemisphere of prosody processing. To our knowledge only one study has so far investigated speech prosody under different processing condition (Dara & Pell, 2006). However, the paradigm used in that study focussed on investigating emotional prosody in patients. RHD patients were impaired in an explicit emotional prosody task, although they were able to implicitly process emotional prosody. Thus, one goal of the present study was to investigate prosody processing in implicit and explicit processing conditions in healthy subjects using non-affective prosodic manipulations.

The second focus of our study was the investigation of cortical interaction effects found for semantic and prosodic speech manipulations. To our knowledge only one previous study has so far focussed on the interaction between prosodic aspects of speech and other speech features, e.g. syntax (Humphries, Love, Swinney, & Hickok, 2005), identifying an interaction in the left superior temporal cortex. The interaction between semantic and prosodic processing has so far only been investigated using electrophysiological methods (Kotz & Paulmann, 2007). Astesano and colleagues, who investigated semantic-prosodic interactions using the same stimulus material as used in our study, reported a P800 for semantic processing which was modulated by prosody processing (Astesano, Besson, & Alter, 2004).

With reference to neuroimaging research on semantic processing, Bookheimer summarized early fMRI-findings suggesting that the cortical network involved in semantic processing is very broad, involving the left inferior frontal gyrus as well as the left fusiform gyrus and middle temporal lobe (Bookheimer, 2002). However, this wide distribution of the semantic processing network seems to depend on the semantic content (Chao, Haxby, & Martin, 1999; Rosazza et al., 2003) and the mode of acquisition (Saffran, Coslett, & Keener, 2003). The involvement of the

posterior middle temporal gyrus has been reported in auditory perceptive tasks such as lexical decision (Friederici, Opitz, & von Cramon, 2000), semantic categorization (Palti, Ben Shachar, Hendler, & Hadar, 2007; Friederici, Ruschemeyer, Hahne, & Fiebach, 2003) as well as picture naming (Indefrey & Levelt, 2004). An involvement of the inferior frontal gyrus, specifically the posterior part, in semantic processing has been repeatedly found. Auditory and visual studies reported a bilateral involvement of the IFG in semantic processing (Chou et al., 2006; Kiehl, Laurens, & Liddle, 2002). Unilateral involvement of the left IFG has been reported in semantic integrative mechanisms after the perception of primed targets (Copland, de Zubicaray, McMahon, & Eastburn, 2007) and in the processing of words as compared to non-words (Wu, Cai, Kochiyama, & Osaka, 2007). We hypothize that an interaction between semantic and prosodic processing could take place in brain areas usually involved in semantic processing (pMTG, BA45) or in brain areas involved in prosodic processing (PT, BA44).

To summarize, the goal of the present study was twofold. First, to investigate the nature of task-modulation on brain areas involved in prosodic processing and, secondly, to investigate the interaction between prosodic and semantic stimulus processing.

**Material and methods**

Participants

16 paid volunteers (7 male, mean age 31) participated in this study. The subjects were native French speakers and none of them had a history of neurological, major medical, psychiatric or hearing disorders. All participants were right handed according to the Annett handedness scale (Annett, 1992). After information about the fMRI procedure was given, written informed consent was obtained from all participants. Two subjects had to be excluded from the study due to technical errors.

Stimulus material and experimental task

Most fMRI-studies on prosodic processing have used manipulated stimulus material which was either filtered, spectrally manipulated or flattened (Meyer et al., 2002; Beaucousin et al., 2007). For the purpose of our study we chose to use a natural prosodic manipulation that can be perceived as prosodically incongruouent but nevertheless contains normal illocutionary prosody. In French, more than in other languages, the illocutionary prosody is clearly defined. We used stimulus material which was formerly used in an ERP study (Astesano et al., 2004) consisting of a total of 240 sentence stimuli. Experimental conditions were constructed according to a 2*2-design including the factors *"prosody"* (congruous and incongruous) and *"semantic"* (congruous and incongruous). In the prosodically congruous condition, half of the sentences were naturally spoken statements and the other half were naturally spoken questions. The prosodic incongruous sentences were produced by cross-splicing the beginning of statements with the end of questions and vice-versa at the syntactic boundary between the noun phrase and the verb phrase. The boundary for the splicing preceded the consonants "p, t, or g" since these consonants are preceded by 50 ms of acoustic noise. By means of this procedure, the time-points at which stimuli were spliced could not be detected in the time course of the sentence. The semantically incongruous condition was constructed by using a verb that does not fit with the noun. For example "La lumière clignotait " ["the light was flickering"] is a semantically congruous sentence while "La toiture traduisait" ["The roof was translating"] is a semantically incongruous sentence. Each of the 4 experimental conditions consisted of 60 stimuli (see Tab. 1). Sentences were built using the same syntactic and phonological pattern (three-syllable subject noun phrases + three-syllable intransitive verbs). For further details and acoustic analysis of the stimuli see Astesano et al. (Astesano et al., 2004). Additionally, 80 empty trials were included in the experiment as it is a standard procedure in fMRI studies. Thus, a total of 320 experimental stimuli were used.

Each subjects had to perform two tasks during the experiment. In the semantic task subjects had to indicate whether the sentences were semantically congruous or incongruous by pressing a button. In the prosody task subjects were asked to pay

attention to the prosody task and decide whether the intonation contour of the sentences was congruous or incongruous. No feedback was given throughout the experiment.

| prosodic manipulation / semantic manipulation | congruous | incongruous |
|---|---|---|
| congruous | n = 60 | n = 60 |
| incongruous | n = 60 | n = 60 |

**Table 1:** Experimental design

*Procedure*

The experiment was conducted in four runs of 80 stimuli each with a short break in between the runs. Half of the subjects performed the semantic task in the first two runs and the prosody task in the third and fourth run. The other subjects performed the task in the reverse order. The order of task performance was counterbalanced within gender. Instructions for the tasks were presented on a screen which subjects viewed from inside the scanner. For the prosodic task these included a visual representation of the melody contour of the sentences. Subjects were explicitly told to pay attention to the pitch progress of the sentences. In a short training session conducted after the instruction, subjects learned to perform the given tasks. Stimuli were presented in pseudorandom order binaurally through headphones. Before each stimulus presentation, a fixation cross was presented for 500 ms.

*Data acquisition*

We implemented a sparse temporal acquisition technique (STA) using an inter-scan interval of 9 seconds acquired 6 seconds after stimulus onset (Zaehle et al., 2007). Data were collected using a Philips Intera 3T whole body MR unit (Philips Medical System Best, The Netherlands) equipped with an eight-channelled Philips SENSE head coil. Functional time series were collected from 16 transverse slices covering the entire perisylvian cortex with a spatial resolution of 2.7 x 2.7 x 4 mm using a Sensitivity Encoded (SENSE) (Pruessmann, Weiger, Scheidegger, & Boesiger,

1999) single-shot gradient-echo planar sequence (acquisition matrix 80 x 80 voxels, SENSE acceleration factor R = 2, FOV = 220 mm, TE = 35 ms and inter-slice gap 2 mm).

Three volumes, consisting of 14 axial slices, were acquired per trial, each with a Tacq=1000 ms, $\theta = 68^{o}$ (decay sampling) and 12 s inter scan interval (ICI). Additionally, we obtained one whole-head EPI volume with 28 transversal slices (Tacq= 3000 ms, $\theta = 82^{o}$) but otherwise the identical scan parameters as for the functional time series to assist in the spatial normalization of the functional data. Furthermore, we collected a standard 3D T1-weighted scan for anatomical reference with 1 x 1 x 1.5 mm spatial resolution (180 axial slices, acquisition matrix 224 x 224 voxels, TE = 2.3 ms, TR = 20 ms, $\theta = 20^{0}$).

**Data analysis**

*Behavioral data*

During the experiment the behavioral performance of 16 subjects was measured in terms of reaction time and response accuracy (percentage of error). Data were corrected for outliers (by trimming >2 std above or below mean value). The behavioral measures, response accuracy and reaction time for all trials per condition, were aggregated per individual and subjected to a three-factorial univariate ANOVA with repeated measures using the within-subject factors of "semantic", "prosody", and "task". Subsequently, a two-factorial ANOVA with repeated measures was performed for each task separately. A one-sampled t-test was performed on response accuracy separately for each task group to test if the performance significantly exceeded chance level.

*fMRI data*

The functional imaging data processing was carried out using MATLAB 7.1 (Mathworks Inc., Natiek, MA, USA) and the software package SPM5 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Functional data were realigned to the first volume, corrected for motion artifacts and normalized into standard stereotactic space (voxel size 2 x 2 x 2 mm, template provided by the Montreal Neurological

Institute (Evans et al., 1992)). For spatial smoothing we applied an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations, performed separately on each voxel (Friston et al., 1995). Eight covariates were calculated for each experimental condition and were convolved with a FIR (Finite Impulse Response) function containing one single basis function and a window length of 9 seconds allowing for no interaction and serial correlation. Four experimental runs were treated as four sessions. The serial autocorrelation of the BOLD time series was modeled using a first-order autoregressive mode. Low-frequency drifts were removed using a temporal high-pass filter (mean cut off 128 sec).

On the whole-head level a partitioned error approach led to the construction of statistical comparisons for contrasts of interest on the individual level using t-tests. Contrast-images were then subjected to a random effects analysis using one-sampled t-tests on the second level. Whole head analyses were thresholded at $T = 3.85$ ($p < 0.001$, uncorrected on voxel-level). Only clusters of the significant size of $28 < k > 30$ ($p < 0.05$, corrected for multiple comparisons on cluster level) were reported on the basis of the correction algorithm implemented in spm5.

Regions of interest (ROI) analyses were performed on core language regions such as the anterior and posterior IFG as well as the PT in order to investigate whether there were interactions between conditions (Hesling, Clement, Bordessoules, & Allard, 2005a; Meyer et al., 2002; Plante, Creusere, & Sabin, 2002; Bookheimer, 2002; Xu et al., 2006). Additionally the pMTG was included (Friederici et al., 2000). The posterior and anterior parts of the inferior frontal gyrus were defined on the basis of cytoarchitectonic probability maps (Amunts et al., 1999; Eickhoff, Schleicher, Zilles, & Amunts, 2006a; Eickhoff, Amunts, Mohlberg, & Zilles, 2006b). The subdivisions of the superior temporal plane (PT) were defined on the basis of macroanatomic landmarks (Steinmetz et al., 1989). For the posterior MTG there is no cytoarchitectonic map available, so it was defined by using the Automatic Anatomic Labeling Atlas (AAL) (Tzourio-Mazoyer et al., 2002), refining the ROI between $-8 > y < 45$. Parameter estimates for ROIs were calculated using marsbar (Brett, Anton, Valabregue, & Polin J-P, 2002) for each

ROI separately. Contrast values were then exported and analyzed with SPSS. Subsequently, four-factorial ANOVAs using the within-subject factors "processing-mode", "prosody", "semantic", and "laterality" were performed. Main effects found in the ROI analysis were not reported in the presence of an interaction with the main effects factor (Henson & Penny, 2005). To resolve three-factorial interactions several two-factorial ANOVAs were performed choosing the factors of interest based on previous literature.

**Results**

*Behavioral results*

A three-factorial ANOVA on the performance level revealed a main effect of "task" ($F_{(1, 14)} = 39.79$, $p = 0.000$) for the performance level. However, one-sampled t-tests revealed that subjects performed significantly above chance in both tasks (Prosody Task: $p < 0.05$, $t_{15} = 2.611$; Semantic Task: $p > 0{,}001$, $t_{15} = 25.144$ ). The three-factorial ANOVA performed on the reaction time revealed a main effect of "task" ($F_{(1, 15)} = 22.22$, $p < 0.001$) and of "prosody" ($F_{(1, 15)} = 6.44$, $p < 0.05$). Additionally, an interaction-effect ("semantic" x "prosody" x "task") was found ($F_{(1, 15)} = 9.08$, $p = 0.01$). Subsequently performed two-factorial ANOVAs for each task separately revealed a significant interaction ("semantic" x "prosody") for the prosody task only ($F_{(1, 15)} = 4.94$, $p < 0.05$). The three-factorial ANOVA performed on the performance rate revealed a main effect of task ($F_{(1, 15)} = 33.10$, $p < 0.001$).

*fMRI results*

The first goal of the study was to demonstrate the effect of task-modulation on prosodic processing. Both explicit and implicit processing was assigned with a network of activity. This included bilateral activity in the superior and middle temporal gyrus, anterior insula as well as activity in the supplementary motor area and left postcentral gyrus for the processing of prosodic congruous as well as prosodic incongruous sentences. However, only the explicit processing condition correlated with activity in the right basal ganglia. Figure 1A and Table 1 display activity found for the two prosodic conditions, separately for both processing tasks as well as collapsed over the factor semantic. An interaction between the task-

manipulation and prosodic stimulus manipulation was found in the right medio-basal temporal cortex, namely the Hippocampus, as indicated in Fig1B (left) and Tab1. Furthermore, in explicit processing the right aIFG was more strongly activated whereas in implicit prosodic processing the left aIFG was more strongly activated. ROI analysis on the anterior inferior frontal gyrus showed a significant two-factorial interaction of "processing mode" and "laterality" ($F_{1, 13} = 6.182$, $p < 0.05$). That is, during explicit semantic processing the left aIFG was activated, while conversely, the right aIFG was more strongly activated during performance of the prosodic task. Paired t-tests revealed significant differences in left hemisphere activation between the two tasks ($t_{13} = -2.807$, $p < 0.05$).

The second goal was to show interactions effects between semantic stimulus processing and prosodic stimulus processing. On the whole-head level this analysis revealed activity in the right middle temporal gyrus as indicated in Fig1B (right) and Tab 1. The ROI analysis performed on the data revealed that there is interaction between semantic and prosodic processing observed in several regions of interest, namely the posterior MTG, pIFG and the PT. Three-factorial interactions of "processing mode" and "semantic" x "prosody" have been found in the posterior middle temporal gyrus (pMTG) ($F_{1, 13} = 19.370$, $p < 0.001$), the posterior inferior frontal gyrus ($F_{1, 13} = 5.820$, $p < 0.05$), as well as in the planum temporale ($F_{1, 13} = 7.030$, $p < 0.05$). The directions of these interactions are indicated in Figure 2 and discussed in the discussion section.

Subsequently performed 2-factorial ANOVAs revealed an interaction of "semanticity" and "prosody" ($F_{1, 13} = 41.284$, $p < 0.001$) for prosodic processing only as well as an interaction between "processing mode" and "semanticity" for prosodic congruous sentences only ($F_{1, 13} = 44.555$, $p < 0.001$). An interaction for "processing mode" and "prosody" was present in semantically congruous ($F_{1, 13} = 15.349$, $p < 0.05$) and semantically incongruous sentences ($F_{1, 13} = 10.783$, $p < 0.05$).Two-factorial ANOVAs on the posterior IFG revealed an interaction between "prosody" and "semanticity" in the prosody task only ($F_{1, 13} = 4.713$, $p < 0.05$) as well as a tendency for an interaction of "processing mode" and "prosody" ($F_{1, 13} = 4.525$, $p = 0.053$) for semantically congruous sentences. A two-factorial

ANOVA on the planum temporale revealed an interaction of "semanticity" and "prosody" ($F_{1, 13} = 6.947$, $p < 0.05$) for prosodic processing only. A two-factorial ANOVA on semantic congruous sentences only revealed a main effect of prosody ($F_{1, 13} = 8.169$, $p < 0.05$) and an effect of processing mode ($F_{1, 13} = 8.680$, $p < 0.05$) when no semantic incongruence was present. In the presence of a semantic incongruence an interaction between prosody and processing mode was observed ($F_{1, 13} = 6.514$, $p < 0.05$).

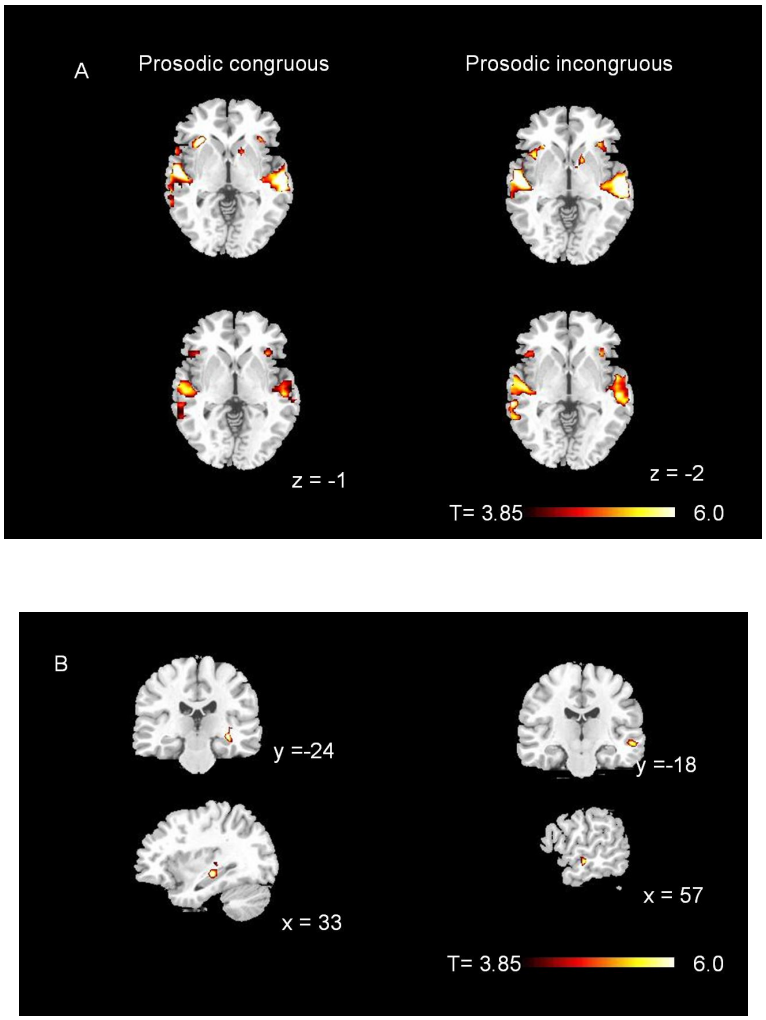



**Figure 1:** The figure shows A) activity for explicit prosodic processing (above) and implicit prosodic (semantic) processing (below) for prosodic congruous and incongruous sentences; B) Interaction effect between prosodic processing and processing mode (left) and between prosodic and semantic processing (right). Activity is displayed neurologically (r=r) with a spatial extent of k = 28 ($p < 0.05$ cluster corrected) superimposed onto an anatomical template.

**Table 2** indicates significant activation clusters on the basis of a whole-head analysis ($p < 0.001$ uncorrected on voxel level and $p < 0.05$ ($k > 27$), cluster corrected). T-values for peak activation and p-values on cluster level are reported.

| **Tab 2** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Explicit processing: Prosodically congruous sentences* | | | | | | | | |
| one-sampled t-test | | | | | | | | |
| | Left hemisphere | | | | Right hemisphere | | | |
| Anatomical description | p- value | T-value | cluster size | x,y,z [mm] | p-value | T-value | cluster size | x,y,z [mm] |
| Middle temporal gyrus | | | | | 0,001 | 11,11 | 884 | 60 -12 -15 |
| Supplementary Motor Area | | | | | 0,001 | 7,63 | 283 | 0 6 48 |
| Insula | | | | | 0,001 | 6,25 | 62 | 36 24 -6 |
| Basla ganglia (Putamen) | | | | | 0,004 | 5,69 | 28 | 15 9 -6 |
| Middle/superior temporal gyrus | 0,001 | 8,88 | 1172 | -54 0 -15 | | | | |
| Postcentral gyrus | 0,001 | 8,12 | 260 | -48 -12 54 | | | | |
| Insula | | | | -27 24 -3 | | | | |
| p-value FDR-.corrected for entire volume on voxel-level, only clusters of p < 0.05 (k = 28), corrected on cluster level are repo | | | | | | | | |
| p- and T-value for peak activation are reported | | | | | | | | |
| *Implicit processing: Prosodically incongruous sentences* | | | | | | | | |
| one-sampled t-test | | | | | | | | |
| | Left hemisphere | | | | Right hemisphere | | | |
| Anatomical description | p- value | T-value | cluster size | x,y,z [mm] | p-value | T-value | cluster size | x,y,z [mm] |
| IFG (pars orbitalis) into temporal pole | | | | | 0,008 | 8,66 | 1031 | 45 21 -18 |
| Postecentral gyrus / inf par lobe | 0,008 | 7,81 | 426 | -45 -27 48 | 0,008 | 7,78 | 663 | 60 -12 -15 |
| Middle temporal gyrus | | | | | | | | |
| middle temporal gyrus | 0,008 | 7,46 | 1218 | -57 -15 -15 | | | | |
| Supplementary motor area | 0,008 | 6,41 | 225 | -3 -6 57 | | | | |
| Thalamus | 0,008 | 6,17 | 77 | -12 -18 6 | | | | |
| Inferior frontal gyrus (pars opercula | 0,008 | 6,36 | 40 | -39 6 27 | | | | |
| p-value FDR-.corrected for entire volume on voxel-level, only clusters of p < 0.05 (k = 30), corrected on cluster level are repo | | | | | | | | |
| p- and T-value for peak activation are reported | | | | | | | | |

| *Explicit processing: Prosodically incongruous sentences* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| one-sampled t-test | | | | | | | | |
| | Left hemisphere | | | | Right hemisphere | | | |
| Anatomical description | p- value | T-value | cluster size | x,y,z [mm] | p-value | T-value | cluster size | x,y,z [mm] |
| Middle temporal gyrus | | | | | 0,001 | 10,45 | 1049 | 60 -12 -15 |
| Supplementary Motor Area | | | | | 0,001 | 10,18 | 536 | 0 -3 57 |
| Basal ganglia (Putamen) | | | | | 0,002 | 6,05 | 107 | 15 9 -6 |
| IFG (orbitalis) extending into Insula | | | | | 0,003 | 5,66 | 99 | 36 27 -3 |
| Middle temporal gyrus | 0,001 | 9,52 | 1045 | -54 6 -21 | | | | |
| Postcentral gyrus | 0,002 | 6,12 | 227 | -48 -24 54 | | | | |
| p-value FDR-.corrected for entire volume on voxel-level, only clusters of p < 0.05 (k = 28), corrected on cluster level are repo | | | | | | | | |
| p- and T-value for peak activation are reported | | | | | | | | |
| *Implicit processing: Prosodically congruous sentences* | | | | | | | | |
| one-sampled t-test | | | | | | | | |
| | Left hemisphere | | | | Right hemisphere | | | |
| Anatomical description | p- value | T-value | cluster size | x,y,z [mm] | p-value | T-value | cluster size | x,y,z [mm] |
| superior temporal gyrus | | | | | 0,002 | 9,33 | 754 | 54 -24 3 |
| Inferior frontal gyrus (pars orbitalis) | | | | | 0,006 | 5,89 | 87 | 45 21 -9 |
| Postcentral gyrus | 0,002 | 9,25 | 1230 | -45 -30 51 | | | | |
| Supplementary motor area | 0,003 | 7,98 | 254 | -6 3 54 | | | | |
| Precentral gyrus | 0,005 | 6,63 | 93 | -39 6 30 | | | | |
| Inferior frontal gyrus (pars orbitalis) | 0,009 | 5,29 | 51 | -48 18 -6 | | | | |
| p-value FDR-.corrected for entire volume on voxel-level, only clusters of p < 0.05 (k = 28), corrected on cluster level are repo | | | | | | | | |
| p- and T-value for peak activation are reported | | | | | | | | |

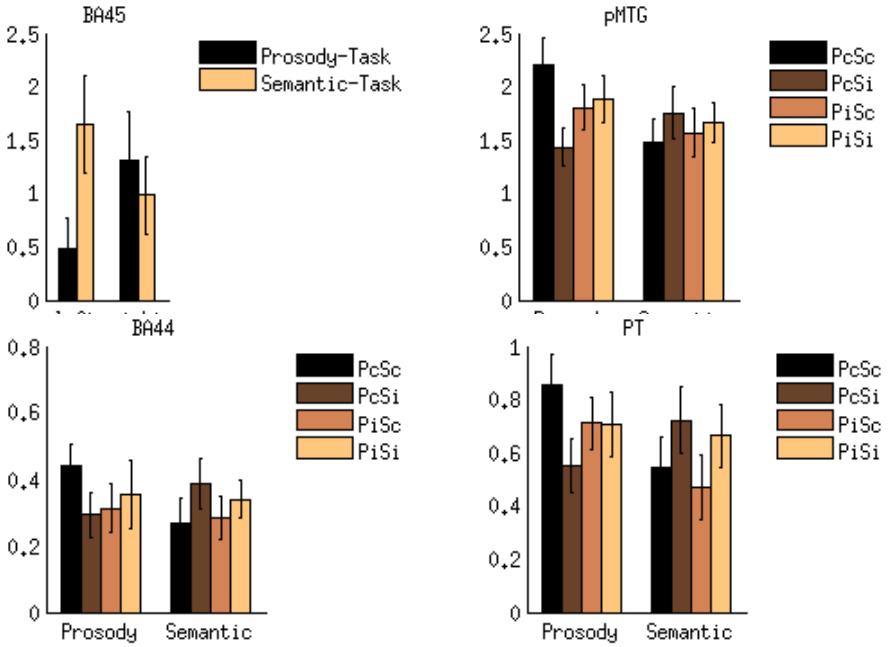



**Figure 2:** Top) Results of ROI analysis for semantically associated ROIs - BA45 and pMTG; Bottom) Results of ROI analysis for prosodically associated ROIs - BA44 and PT; "Prosody" = prosodic task, "Semantic" = semantic task, "Pc" = prosodically congruous, "Pi" = prosodically incongruous, "Sc" = semantically congruous, "Si" = semantically incongruous, "left-H" = left hemisphere, "right-H" = right hemisphere; the x-axis represents the mean percent signal change as calculated by marsbar on the basis of beta values.

## Discussion

Generally, our data replicates former findings on prosodic processing that show its association with activity in the superior temporal gyrus and the anterior insulae. In addition, the explicit processing of prosodic stimuli correlated with activity in the basal ganglia (Putamen, neostriatum), which is in line with earlier reported findings (Meyer et al., 2002; Pell & Leonard, 2003; Kotz et al., 2003).

### *Task modulation on stimulus processing*

We report a new finding that the processing task modulates stimulus responsiveness in the posterior middle temporal gyrus. The right middle temporal gyrus has repeatedly been found in semantic processing (Jung-Beeman, 2005). Task modulation is evident in the interaction between the processing-mode and semantic in the prosodic congruent stimuli on the one hand, and on the other hand the interaction between processing mode and prosody in semantically congruent

stimuli. That is, across the same stimulus type depending on the task performed, the posterior middle temporal gyrus responds more strongly to a specific stimulus characteristic, namely to congruency vs. incongruency. This observation strikes an often overlooked issue in brain imaging research. Stimuli which violate the stimulus characteristic under investigation are often constructed and compared to stimuli which do not violate this stimulus characteristic (Friederici et al., 2003; Meyer, Friederici, & von Cramon, 2000; Newman & Twieg, 2001). However, the opposite approach, comparing the intact condition against the violated condition in order to isolate the process of investigation, has been taken as well (Xiao et al., 2005). On the basis of the present findings one might argue that the efficiency of each approach depends on the specific task implemented. It is possible that violation conditions are of particular interest in explicit processing and that non-violation conditions are of interest in implicit processing. However, this interpretation is still a matter of speculation and needs to be investigated in more detail. From our observation we conclude that our date provide evidence of a strong top-down modulation on the auditory association cortex (Scheich et al., 2007; Geiser, Zaehle, Jancke, & Meyer, 2008), and, furthermore, that we must strongly focus especially on the processing condition when comparing fMRI studies

In terms of the effect of task on the processing of prosodic stimulus we surprisingly found a significant interaction effect between task and prosodic manipulation in the right hippocampus (cornu ammonis). The medial temporal lobe (MTL), namely the parahippocampal gyrus, is essential in memory related recognition in case of novelty, familiarity or recency (Yassa & Stark, 2008). In our data this brain area responds more strongly to prosodic incongruence during the performance of a speech prosody task while its activity is higher in prosodic congruent stimuli during the performance of semantic task. Since the task, namely the prosody task, was difficult and had to be trained prior to performance of the experiment, we are tempted to interpret this finding as a case of identification of previously trained stimuli in a manner similar to the finding recently reported on melodic stimuli by Watanabe et al. for the right hippocampus (Watanabe, Yagishita, & Kikyo, 2008). It is interesting to note that during the performance of the semantic task this activity

reversed in that the right hippocampus responded more strongly to prosodically congruent stimuli. It appears that while subjects focus their attention on other stimulus characteristics, the hippocampus nevertheless responds to the familiarity of the prosodic congruent stimulus (Gonsalves, Kahn, Curran, Norman, & Wagner, 2005).

We observed a differentially lateralized involvement in the anterior inferior frontal cortex depending on the task performed. The left hemispheric aIFG showed significantly stronger activation in the semantic task than in the prosodic task. The right hemisphere showed the opposite pattern. This task related functional lateralization, which is independent of the prosodic or semantic stimulus characteristics, is in accordance with the well established opinion that this brain area serves a specific function in explicit stimulus processing. The involvement of the left posterior IFG in semantic processing is well documented. It has formerly been found to be involved in semantic categorization as well as semantic decision tasks (Chou et al., 2006; Copland et al., 2007; Palti et al., 2007; Meyer et al., 2004; Bookheimer, 2002; Poldrack et al., 1999). In contrast the right inferior frontal gyrus has been found to be involved in linguistic as well as non-linguistic pitch processing (Liu et al., 2006; Humphries, Willard, Buchsbaum, & Hickok, 2001; Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2004; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Thus, according to our data, activity in the inferior frontal gyrus is not primarily modulated by the specific stimulus characteristic but rather comes into play as soon as explicit processing is required.

*Interaction between semantic and prosodic processing*

The second aim of the analysis was to investigate the interaction between semantic and prosodic stimulus processing. Whole head analysis revealed an interaction between the two stimulus manipulations in the right middle temporal gyrus, in that this brain area is more strongly activated when both stimulus characteristics were either congruous or incongruous. In addition, the results of the ROI analysis revealed an interaction in the *auditory association cortex,* specifically in attentive prosody processing. The pMTG showed greater activation for semantically congruous than for incongruous sentences while in the presence of a prosodic

incongruence, this effect was reversed. Interestingly, both the frontal brain area BA44 in the inferior frontal gyrus and the planum temporale displayed the same effect. Thus, we conclude that the middle temporal gyrus is sensitive to prosodic (Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003) as well as semantic (Friederici et al., 2000) stimulus characteristics and that these two stimulus characteristics mutually interact.

**Conclusion**

We identified auditory association cortex areas, namely the PT as well as the MTG as brain areas in which an interaction between semantic and prosodic processing can take place. In addition, we observed that the responsiveness to congruent vs. incongruent stimulus characteristics in the pMTG depends on the task performed. Furthermore, our results are in concurrence with former findings on the planum temporale involvement in speech prosody as well as the involvement of the posterior middle temporal gyrus and the left anterior inferior frontal gyrus in semantic processing.

Reference List

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. J.Comp Neurol., 412, 319-341.

Annett, M. (1992). 5 Tests of Hand Skill. Cortex, 28, 583-600.

Astesano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. Brain Res.Cogn Brain Res., 18, 172-184.

Baum, S. R. & Dwivedi, V. D. (2003). Sensitivity to prosodic structure in left- and right-hemisphere-damaged individuals. Brain Lang, 87, 278-289.

Beaucousin, V., Lacheret, A., Turbelin, M. R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. Cereb.Cortex, 17, 339-352.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu.Rev.Neurosci., 25, 151-188.

Brechmann, A. & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cerebral Cortex, 15, 578-587.

Brett, M., Anton, J.-L., Valabregue, R., & Polin J-P (2002). Region of interest analysis using an SPM toolbox. In Presented at the 8th International Conference on Functinal Mapping of the Human Brain. Sendai, Japan.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat.Neurosci., 2, 913-919.

Chou, T. L., Booth, J. R., Burman, D. D., Bitan, T., Bigio, J. D., Lu, D. et al. (2006). Developmental changes in the neural correlates of semantic processing. Neuroimage., 29, 1141-1149.

Copland, D. A., de Zubicaray, G. I., McMahon, K., & Eastburn, M. (2007). Neural correlates of semantic priming for ambiguous words: an event-related fMRI study. Brain Res., 1131, 163-172.

Dara, Ch. & Pell, M. D. (2006). Effects of right-hemisphere damage on explicit and iomplicit processing of emotional prosody. Brain and Language, 99, 8-219.

Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., & Caplan, D. (2004). Question/statement judgments: an fMRI study of intonation processing. Hum.Brain Mapp., 23, 85-98.

Eickhoff, S. B., Amunts, K., Mohlberg, H., & Zilles, K. (2006b). The human parietal operculum. II. Stereotaxic maps and correlation with functional imaging results. Cereb.Cortex, 16, 268-279.

Eickhoff, S. B., Schleicher, A., Zilles, K., & Amunts, K. (2006a). The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. Cereb.Cortex, 16, 254-267.

Evans, A. C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W. et al. (1992). Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage., 1, 43-53.

Friederici, A. D. & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang, 89, 267-276.

Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. Cereb.Cortex, 10, 698-705.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb.Cortex, 13, 170-177.

Friston, K. J., Holmes, A. P., Worsley, K. P., Poline, J. B., Frith, C. D., & Frackowiak, R. S. (1995). Statistical parameter maps in functional imaging: A general linear approach. Hum.Brain Mapp., 2, 189-210.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L. et al. (2003). Temporal integration of speech prosody is shaped by language experience: an fMRI study. Brain Lang, 84, 318-336.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y. et al. (2004). Hemispheric roles in the perception of speech prosody. Neuroimage., 23, 344-357.

Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. J.Cogn Neurosci., 20, 541-552.

Gonsalves, B. D., Kahn, I., Curran, T., Norman, K. A., & Wagner, A. D. (2005). Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. Neuron, 47, 751-761.

Henson, R. N. A. & Penny, W. D. (2005). ANOVAs and SPM Wellcome Department of Imaging Neuroscience.

Hesling, I., Clement, S., Bordessoules, M., & Allard, M. (2005). Cerebral mechanisms of prosodic integration: evidence from connected speech. Neuroimage., 24, 937-947.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. Neurosci.Res., 54, 276-280.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang, 76, 227-252.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Hum.Brain Mapp., 26, 128-138.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. Neuroreport, 12, 1749-1752.

Indefrey, P. & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. Cognition, 92, 101-144.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends Cogn Sci., 9, 512-518.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: an event-related fMRI study of semantic processing. Neuroimage, 17, 842-850.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang, 86, 366-376.

Kotz, S. A. & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. Brain Res., 1151, 107-118.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. Hum.Brain Mapp., 24, 11-20.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K. et al. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. Neuroimage., 29, 515-523.

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp., 17, 73-88.

Meyer, M., Friederici, A. D., & von Cramon, D. Y. (2000). Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Brain Res.Cogn Brain Res., 9, 19-33.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang, 89, 277-289.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia, 41, 1410-1421.

Newman, S. D. & Twieg, D. (2001). Differences in auditory processing of words and pseudowords: an fMRI study. Hum.Brain Mapp., 14, 39-47.

Palti, D., Ben Shachar, M., Hendler, T., & Hadar, U. (2007). Neural correlates of semantic and morphological processing of Hebrew nouns and verbs. Hum.Brain Mapp., 28, 303-314.

Pell, M. D. (2005). Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang.

Pell, M. D. & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. Cogn Affect.Behav.Neurosci., 3, 275-288.

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: activation interacts with task demands. Neuroimage., 17, 401-410.

Plante, E., Holland, S. K., & Schmithorst, V. J. (2006). Prosodic processing by children: an fMRI study. Brain Lang, 97, 332-342.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K. et al. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. Brain Res.Cogn Brain Res., 4, 231-242.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage, 10, 15-35.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. Magnetic Resonance in Medicine, 42, 952-962.

Rodway, P. & Schepman, A. (2007). Valence specific laterality effects in prosody: expectancy account and the effects of morphed prosody and stimulus lead. Brain Cogn, 63, 31-41.

Rosazza, C., Imbornone, E., Zorzi, M., Farina, E., Chiavari, L., & Cappa, S. F. (2003). The heterogeneity of category-specific semantic disorders: evidence from a new case. Neurocase., 9, 189-202.

Saffran, E. M., Coslett, H. B., & Keener, M. T. (2003). Differences in word associations to pictures and words. Neuropsychologia, 41, 1541-1546.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & Ohl, FW. (2007). The Cognitive Auditory Cortex: Task-Specificity of Stmulus REpresentations. Hearing Research, 229, 213-24.

Steinmetz, H., Rademacher, J., Huang, Y. X., Hefter, H., Zilles, K., Thron, A. et al. (1989). Cerebral asymmetry: MR planimetry of the human planum temporale. J.Comput.Assist.Tomogr., 13, 996-1005.

Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev., 43, 231-246.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage., 15, 273-289.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. J.Cogn Neurosci., 15, 1019-1027.

Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A., & Obrig, H. (2006). The processing of prosody: Evidence of interhemipsheric specialization at the age of four. Neuroimage, 34, 416-425.

Watanabe, T., Yagishita, S., & Kikyo, H. (2008). Memory of music: roles of right hippocampus and left inferior frontal gyrus. Neuroimage, 39, 483-491.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog.Brain Res., 156, 249-268.

Wu, J., Cai, C., Kochiyama, T., & Osaka, K. (2007). Function segregation in the left inferior frontal gyrus: a listening functional magnetic resonance imaging study. Neuroreport, 18, 127-131.

Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X. et al. (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: an event-related fMRI study on auditory lexical decision. Hum.Brain Mapp., 25, 212-221.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidizic, M., Tong, Y. et al. (2006). aCtivation of the left planum temporale in ptich processing is shaped by language experience. Human Brain Mapping, 27, 173-83.

Yassa, M. A. & Stark, C. E. (2008). Multiple signals of recognition memory in the medial temporal lobe. Hippocampus.

Zaehle, T., Schmidt, C. F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P. et al. (2007). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. Neuroimage, (epub ahead of print).

Zatorre, R. J. (2001). Neural specializations for tonal processing. Biological Foundations of Music, 930, 193-210.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A. et al. (2005). Speech recognition with amplitude and frequency modulations. Proc.Natl.Acad.Sci.U.S.A, 102, 2293-2298.

## Experiment III: Electrophysiological correlates of meter and rhythm processing in music perception

Authors: Geiser E., Ziegler, E., Jancke L. & Meyer M.
Published in Cortex

**Abstract**

The two main characteristics of temporal structuring in music are *meter* and *rhythm*. The present experiment investigated the event-related potentials *(ERP)* of these two structural elements with a focus on differential effects of attended and unattended processing. The stimulus material consisted of an *auditory* rhythm presented repetitively to subjects in which metrical and rhythmical changes as well as pitch changes were inserted. Subjects were to detect and categorize either temporal changes (attended condition) or pitch changes (unattended condition). Furthermore, we compared a group of long-term *trained* subjects (*musicians*) to non-musicians. As expected, behavioural data revealed that trained subjects performed significantly better than untrained subjects. This effect was mainly due to the better detection of the meter deviants. Rhythm as well as meter changes elicited an early negative deflection compared to standard tones in the attended processing condition, while in the unattended processing condition only the rhythm change elicited this negative deflection. Both effects were found across all experimental subjects with no difference between the two groups. Thus, our data suggest that meter and rhythm perception could differ with respect to the time course of processing and lend credence to the notion of different neurophysiological processes underlying the auditory perception of rhythm and meter in music. Furthermore, the data indicate that non-musicians are as proficient as musicians when it comes to rhythm perception, suggesting that correct rhythm perception is crucial not only for musicians but for every individual.

## Introduction

Melody and harmonic changes are commonly perceived as the characteristic elements of music. However, just as essential to the perception of music is its temporal structure, that is, meter and rhythm. A fast rhythm can induce a feeling of excitement or slower rhythms can alleviate feelings of tension. On the other hand, spontaneous body movements induced by music are most often related to the perceived meter and movement can in fact enhance meter perception in auditory rhythm (Phillips-Silver & Trainor, 2007). In sum, temporal structure crucially constitutes the specific character of a musical sequence. The present study aims to investigate the correlate of meter and rhythm perception using event-related potentials.

There have been several descriptive approaches taken to define the constituting elements of rhythm and meter. Cooper & Meyer (1960) defined rhythm as the temporal relation between one or several unaccented tones and an accented tone. Schachter describes rhythm as "patterns of durations, emphases and groupings" that do not arise from pitch structure (Schachter, 1986). Lerdahl und Jackendoff (1983) use the term "grouping structure" to mean a hierarchical segmentation of music in motive, phrases and sections based on tone groups belonging to each other according to their temporal relation.

While rhythm is related to the duration of tones, meter is related to the relative accents of tones within their context. Lerdahl und Jackendoff describe meter as psychological extrapolation of isochronously defined tones perceived through metrical accents (Lerdahl et al., 1983). Similarly, Cooper & Meyer (1960) defined meter as the number of pulses between more or less regularly accented tones. One of these authors found that a perceived meter is relatively resistant to deviants, allowing listeners to perceive a syncope within the context of a given meter (Meyer, 1956). This paved the way to Cooper & Meyer focussing on the psychological characteristic of the meter percept, describing it as a Gestalt phenomenon.

Most authors agree that the emergence of a meter percept depends on the perception of accent patterns. However, accents can arise from different characteristics of tones within their musical context. Recent experimental research has tended to focus on temporal, dynamic, and melodic accents. A temporal accent may appear when a tone

is relatively isolated, the second of a two-tone cluster, or the initial or final note of a cluster of three or more notes. A dynamic accent is defined as a relative change of intensity on the accentuated tone (Drake & Palmer, 1993), while a melodic accent consists of a higher pitch or a "point of change within the melody" (Trehub & Hannon, 2006). Recent studies showed that the perception of meter is induced by both melodic and temporal accents (Hannon, Snyder, Eerola, & Krumhansl, 2004, Povel & Essens, 1985) and that in the absence of melodic accents dynamic accents become important for meter perception (Toivainen & Eerola, 2004).

A cognitive approach describing the perception of music has postulated modular or domain-specific processing, hereby assuming different processing mechanisms underlying the perception of structural elements. It has been generally formulated for the processing of pitch as compared to the processing of temporal structure (Lerdahl et al., 1983) as well as in association with specific temporal structuring, that is, for meter and rhythm. The latter was embedded in the "two component model" by Lerdahl and Jackendoff (Lehrdahl & Jackendoff, 1981).
The advancement of neurophysiological research methods has enhanced this cognitive approach and lea=d to the suggestion of different neural underpinnings related to the specific processing modules (Peretz, 2006). The neurofunctional differentiation of pitch and temporal processing has found support in a number of lesion and neurofunctional studies and associate temporal processing generally to brain areas of the left hemisphere (Midorikawa, Kawamura, & Kezuka, 2003; Piccirilli, Sciarma, & Luzzi, 2000; Peretz & Kolinsky, 1993; Murayama, Kashiwagi, Kashiwagi, & Mimura, 2004). Empirical evidence in support of the "two-component model" of temporal processing is however still sparse and relates mostly to behavioural research (Essens, 1986; Peretz & Zatorre, 2005; Ibbotson & Morton, 1981; Liegeois-Chauvel et al., 1998). Evidence from lesion studies is ambiguous. Some authors have associated meter reproduction with the right hemisphere (Penhune, Zatorre, & Feindel, 1999; Wilson, Pressing, & Wales, 2002), while rhythm discrimination was associated with the left hemisphere (Di Pietro, Laganaro, Leemann, & Schnider, 2004). In contrast, a study on healthy subjects postulated that memorization of metric rhythm relies on left premotor and parietal cortex areas,

while right prefrontal, premotor and parietal cortical areas engage in nonmetrical rhythm processing (Sakai et al., 1999). A similar finding has been observed in a speech production experiment in which the reproduction of rhythmical as compared with isochronous syllables activated fronto-temporal brain areas of the right hemisphere (Riecker, Wildgruber, Dogil, Grodd, & Ackermann, 2002). Thus, present neurofunctional evidence about the different neural underpinnings of meter and rhythm processing is inconclusive.

Due to its excellent temporal resolution, we used electrophysiological measures to elucidate the early time course of brain responses during on-line perception of meter and rhythm and to find evidence for or against functional differentiation between rhythm and meter processing. To our knowledge, there are only two studies that recently investigated auditory temporal processing. One study investigated temporal auditory processing in an attended processing task, using DC-potentials of EEG signals (Kuck, Grossbach, Bangert, & Altenmuller, 2003). Subjects had to perform a same-different task while listening to two subsequent acoustic time sequences. The stimuli differed with respect to either meter or rhythm. Processing of both meter and rhythm produced sustained cortical activation over bilateral frontal and temporal brain regions. Thus, this study speaks against a neurofunctional difference between rhythm and meter processing. However, we believe that the reported activation in the study by Kuck et al. is indeed more related to the comparison of temporal patterns than to the perception of metrical and rhythmical changes as the authors themselves suggest. Especially in the domain of metrical changes, we posit that the comparison of two sequences requires cognitive processes in addition to the mere perception of meter. In a magnetencephalographic study, Vuust and colleagues showed an N100m (the magnet-encephalographic equivalent of the N1, 100-150 ms post stimulus onset) related to the detection of metric deviants, which differ in amplitude depending on the acuteness (sharpness) of metrical change (Vuust et al., 2005). The early negativity was more pronounced over the left hemisphere for experienced listeners and over the right hemisphere for naive listeners.

On the basis of above mentioned findings we assumed that the perception of temporal deviants is reflected by an ERP in the latency window of an endogenous component, that is, > 100 ms after the onset of the deviant tone. Thus, we constructed

our stimulus material with the specific aim of analyzing this latency window. Our study investigated processing differences between auditory rhythm and meter perception by presenting subjects with changes in either of the two categories. Our study is novel in several respects. First, subjects are continuously presented with a metrical rhythm. By means of this procedure we are able to induce a strong feeling of meter prior to a deviant stimulus. Secondly, we introduced "processing mode", that is, an attended and an unattended processing condition, as an additional experimental factor. The aim of this procedure was to clarify whether correlates of meter and rhythm changes are differentially affected by the processing mode. Finally, we investigated musically long-term trained and untrained individuals under the assumption that long-term training would impact differentially on meter and rhythm and that such training effects could be a strong indicator to differential processing mechanisms.

**Procedure**

*Stimulus Material / Experimental Conditions*

The basis for the stimulus material consisted of an auditory rhythm in ¾ meter on one note ("g'"), which was presented continuously in tempo 100 Hz (=0.6 s per quarter note). The rhythm was devised in such a way that we had a temporal accent at the beginning of each bar. Therefore, the rhythm consisted of a quarter note at the beginning of each bar followed by four eighth notes (Fig. 1).

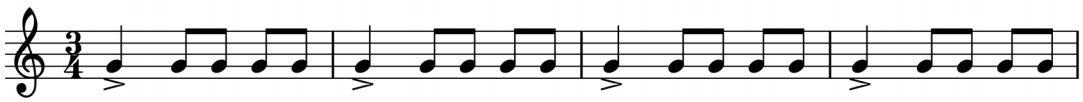


**Figure 1:** Basis of the stimulus material

The stimulus material included *temporal deviants* that is, two different metric and two different rhythmic changes (Fig.2). The metric change consisted either of an insertion of one eight note producing a bar in ⅞ instead of 3/4 meter (condition M7, n=82) or a removal of an eighth note producing bar in ⅝ instead of ¾ meter (condition M5, n=82). The rhythmic change consisted either of an insertion of two

sixteenth notes producing a group of four sixteenth notes on the third beat (condition R4, n=82) or an insertion of one sixteenth producing a group of an eighth and two sixteenths on the third beat (condition R2, n=82).

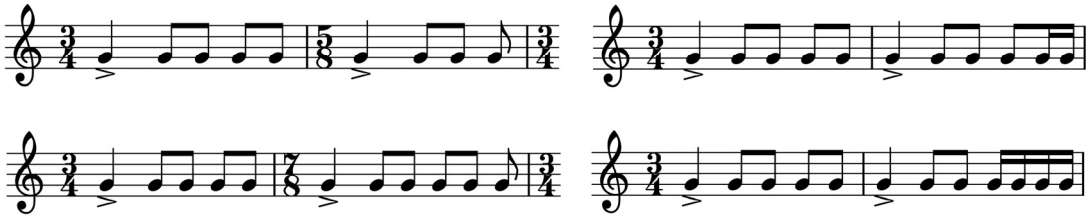


**Figure 2:** Two different types of metric changes (left) and two different types of rhythmic changes (right)

Additionally, pitch manipulations consisting of an "e'" or a "b' flat" on the first ⅛ of the second beat in a bar were included in the stimulus material and balanced over temporal conditions (Fig. 3).

The stimulus material was designed in such a way that there were two, three, or four bars (4:2:1) between deviant bar and this number of basis bars was randomized over experimental conditions. There was a total number of 82 deviant bars per condition.

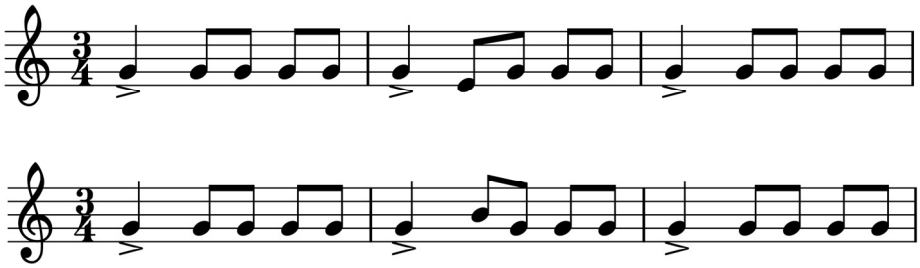


**Figure 3:** Two pitch manipulations

Stimuli were constructed using the open-source LilyPond software (http://lilypond.org/web/) and exported as midi-files. Then the rhythms were synthesized using a piano-sound on one note (g') using "fruity-loop" software (http://www.flstudio.com/) and exported as .wav files. A dynamic accent of 30 percent above the general intensity was induced on the first beat of every bar to reinforce the perceived meter (Hannon et al., 2004). This accent is indicated by a ">" in Figures 1, 2, and 3.

**Task / Experimental design**

All subjects were to perform two tasks in two subsequent experimental blocks. In the first block, subjects had to detect and categorize pitch changes. That is, they had to press the right of two buttons when perceiving an "e'" and the left button when perceiving a "b' flat". This condition is referred to as "unattended" processing condition, since rhythm deviants are perceived inattentively. In the second block, subjects had to detect and categorize temporal changes. That is, they had to press the right button when perceiving a meter change and the left button when perceiving a rhythm change (attended condition). Therefore, the resulting experimental design was a three factorial design with one between-subject factor ("group"), and two within-subject factors ("processing mode" and "rhythmicity") as indicated in Table 1.

**Table 1:** Experimental design

| | | attended | unattended |
|---|---|---|---|
| musicians | meter | M5 (n = 82) | M5 (n = 82) |
| | rhythm | R2 (n = 82) | R2 (n = 82) |
| nonmusicians | meter | M5 (n = 82) | M5 (n = 82) |
| | rhythm | R2 (n = 82) | R2 (n = 82) |

Two within-subject factors "processing mode" (attended-unattended) and "temporal manipulation" (rhythm – meter). One between-subject factor "group" (musicians – non-musicians).

**Experimental groups / Subjects**

Two groups of male subjects, a musically trained group (n=15, mean age = 29.6, std = 4.9) and a musically untrained group (n=15, mean age = 31.1, std = 6.2), participated in the experiment. Untrained subjects did not play an instrument and had no musical training at all. Trained subjects were students at the conservatory or musicians. To substantiate the group categorization, an auditory test (Gordon, 1989) was performed. A one-way independent-sampled t-tests performed on these data revealed significant group differences between the two groups for the rhythm task ($t_{(28)} = 3.236$; $p < 0.005$) as well as for the tonal task ($t_{(28)} = 3.322$; $p = 0.001$).

Musicians performed significantly better than non-musicians in both the rhythm and the tonal task. No subjects had to be excluded on the basis of their rhythm performance with respect to the group performance or with respect to the standardized norm provided by the authors of the test. However, one subject of the musician group had to be excluded from the analysis after EEG-artefact correction. Two further subjects, one from each group, had to be excluded from the analysis due to percent of correct answers below 2 standard deviations of the mean group performance. Thus, we included a total of 13 musicians and 14 non-musicians in our analysis. All subjects were right handed according to the Annett handedness scale (Annett, 1992), with no history of neurological, major medical or psychiatric disorders and with normal audiological status. Furthermore, subjects gave their consent in accordance with guidelines of the Ethical Committee of the local Medical Faculty.

**Experimental procedure**

After receiving written and oral instruction, subjects were made familiar with the task. During the experiment a total number of 328 deviant bars were pseudo-randomly presented using Hifi-headphones. The total number of stimuli was presented twice, each block lasting 32 minutes. The subjects first performed the pitch detection task in the first block and, after a short break, the attended task (temporal task). The stimuli were presented via headphones in a sound and light attenuated room. Stimulus presentation and answer registration was done using Presentation Software (http://www.neurobs.com/).

**EEG-Recordings**

EEGs were recorded continuously using a QuickAmp EEG amplifier (Brainp Products GmbH, 2004) and applying a bandpass filter of 0.1-100 Hz, a notch-filter of 50 Hz. The EEG signal was continuously sampled and digitalized at 500 Hz. EEGs were recorded from 30 electrode sites using a subset of the 10-10 system (Chatrian, lettich, & Nelson, 1988) provided by Easy Cap. Vertical and horizontal EOGs were recorded from 2 bipolar electrodes placed on the infra-orbital ridges of the left and the right eye and the outer canthi of the two eyes. All leads were referenced to a

common average reference. The ground electrode was positioned fronto-centrally (AFz). EEG data were scanned off-line for eye movements and other artefacts. When the maximum absolute voltage difference within an interval of 200 ms exceeded 100 μV or the electrical activity generally exceeded ± 100 μV, the signal was excluded from analysis in the time range of 100 ms before and after the rhythmic and metric change respectively.

**Analysis**

*Behavioural Data*

We analysed percent correct answers and reaction time for the two conditions. Since we adopted a two-alternative forced choice task, the percentage of correct answers calculated across the meter deviants as well as the rhythm deviants was taken as a measure for task performance. As a measure for discrimination rate, d' was calculated (MacMillan N.A., 1991). A repeated measures (2x2) ANOVA using the within-subject factor "rhythmicity" (meter and rhythm) and the between-subject factor "group" (musicians and non-musicians) was performed for the task performance. Furthermore, a one-way independent sample t-test was performed on the discrimination rate.

**EEG Data**

EEG data were analyzed using Brainvision Analyzer 1.0 (Brain Products GmbH, Munich). Figure 4 illustrates the time point within the stream of stimuli from which we calculated the event- related potentials for each stimulus condition. ERPs for deviants were calculated for both processing conditions (attended/unattended) time-locked to the same note. For the meter deviants the ERPs were calculated on the first quarter tone of the bar following the ⅝ bar (condition M5) and on the last eighth note of the ⅞ bar (condition M7) respectively. For the rhythm changes the ERPs were calculated on the second sixteenth note of the 4-sixteenth group The ERPs of the rhythm and meter deviants were compared to ERPs evoked by tones which do not constitute a deviant (n = 82). This non-deviant tone was different for each experimental condition. The non-deviant tones had to show the same intensity as the deviant tones and had to be preceded by a tone of the same length as the antecedent

tone of the deviants. These non-deviants tones are indicated as "control" in Figure 4. The ratio between deviants and standards was 1:1 in the rhythm condition and approximately 1:3 in the meter condition. (condition R2) and on the second $^{1}/_{16}$ tone of the $^{2}/_{16}$ group (condition R4).

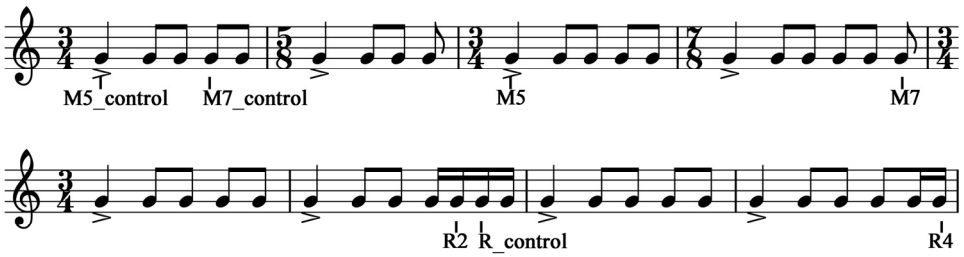



**Figure 4:** Time point within the stimulus from which the event-related potentials (ERPs) were calculated for each stimulus condition. For the meter changes the ERPs were calculated on the first ¼ note of the bar following the 5/8 bar (condition M5) and on the last 1/8 note of the 7/8 bar (condition M7) respectively. For the rhythm changes the ERPs were calculated on the second 1/16 of 4 /16 (condition R2) and on the second 1/16 of the 2/16 group (condition R4). Control tones for each experimental condition are indicated likewise.

The continuous EEG data were segmented off-line into epochs of 800 ms starting 100 ms before the stimulus of the specific experimental condition. After artefact correction a mean of 65.4 deviants per condition, 71.75 rhythm standards, and 195.5 meter standard segments was analyzed per condition and subject. The segments were baselined to the pre-stimulus interval (-100 to 0 ms) and averaged according to stimulus type and processing condition separately for each experimental group. For the attended processing condition only trials which were correctly identified within 3600 ms (3 bars) after the deviant were analyzed.

Analysis of behavioural data revealed, that the performance was better for M5 and R2 as compared to the respective other experimental condition and that this effect was present in both experimental groups (see "stimulus material"). Therefore we restricted our analysis of temporal deviants to these two conditions. The N150 peak was determined as the absolute voltage minimum in the time window between 90 and 160 ms after stimulus onset for the rhythm condition and 110-160 ms after stimulus onset for the meter condition occurring at electrode Cz. R2, M5, and their specific control conditions were separately subjected to repeated measures ANOVA using the between-subject factor "group" (musicians – non musicians), the within-

subject factor "processing mode" (unattended –attended), and "rhythmicity" (R2 – R2 control) / "metricity" (M5 – M5 control), respectively. Interactions in this 3-factorial ANOVA were analyzed by three subsequent 2-factorial ANOVAs.

## Results

*Behavioural Results*

One-way independent sample t-test to identify group differences performed on the discrimination rate revealed significant group differences ($t_{(28)} = 3.126$; $p < 0.05$). The repeated measures ANOVA on the percent of correct answers (Tab 2) using the within-subject factor "rhythmicity" (meter and rhythm) and the between-subject factor "group" (musicians and non-musicians) revealed a main effect of rhythmicity ($F_{(1,26)} = 19.948$; $p < 0.005$ and an interaction "group x rhythmicity" ($F_{(1,26)} = 19.648$ $p < 0.05$) as well as a main effect of group ($F_{(1,26)} = 1379.6$, $p < 0.05$). Two-way independent sample t-tests to resolve the interaction revealed, that the group difference was more pronounced on the meter condition ($t_{(26)} = 3.304$; $p < 0.05$) as compared to the rhythm condition ($t_{(26)} = 2.305$; $p < 0.05$). That is, untrained subjects performed worse on the meter as compared to the rhythm condition. This difference was not found in the group of trained subjects.

**Table 2:** Percent correct answers for musicians and non-musicians

| Meter | Rhythm | M5 | M7 | R2 | R4 | |
|---|---|---|---|---|---|---|
| 89.6 (20.9) | 94.3 (5.6) | 91.1 (17.4) | 88.2 (24.6) | 95.6 (5.1) | 93.0 (6.8) | musicians |
| 75.6 (21.5) | 86.1 (12.8) | 77.6 (17.8) | 73.7 (27.5) | 91.9 (8.0) | 80.4 (21.7) | nonmusicians |

(left) collapsed over meter and rhythm conditions (right) separately for each condition (standard deviation in brackets).

## EEG Results

Figure 5 shows the ERP waveforms to the deviant tones and to the control tones. Red lines illustrate the ERPs to the attended processing condition and blue lines illustrate the ERPs to the unattended processing condition. Figure 5 (left) visualizes the ERPs elicited by the meter for all subjects collapsed over both groups and Figure 5 (right) the ERPs elicited by the rhythm condition for all subjects regardless of group. Results indicate an early negative deflection peaking at 150 ms on average in the

rhythm condition. The meter condition elicited a negativity in the same latency window in attended processing only. Thus, a difference between the two temporal manipulations is reflected in a differential effect of the processing mode on the early negative deflection.

Subsequently we report the results revealed by the three-factorial ANOVA (rhythmicity/metricity x processing mode x group) calculated on the ERP-amplitude separately for rhythm and for meter manipulations. The ANOVA calculated on the ERPs related to *rhythm manipulation* revealed a main effect of "rhythmicity" ($F_{(1, 25)} = 83.493$, $p < 0.001$). No effect of group or processing mode was observed. That is, rhythm deviants revealed significantly higher amplitudes as compared with control tones in both groups and across experimental conditions. Likewise, the ANOVA calculated on the ERPs related to *meter manipulation* revealed a main effect of "metricity" ($F_{(1, 25)} = 38{,}553$, $p < 0.001$) as well as an interaction between "metricity" and "processing mode" ($F_{(1, 25)} = 33.316$, $p < 0.001$). No main effect of group was observed. The interaction was due to the significant effect of "metricity" within the attended processing condition only ($F_{(1, 25)} = 53.096$, $p < 0.001$). There were higher amplitudes observed for the meter deviant as compared with the control tone in the attended processing condition. The amplitude related to the meter deviant in the unattended processing condition did not significantly differ from the control tone.

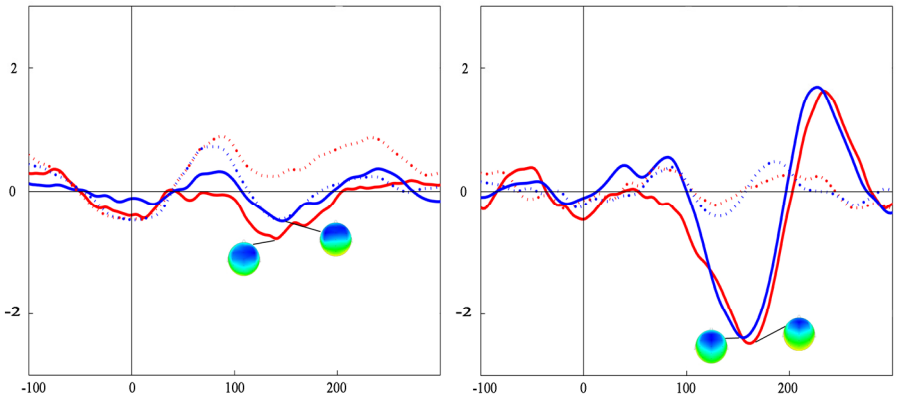



**Figure 5:** ERPs related to deviant and standard stimuli in the explicit and implicit processing condition. (5 left) meter processing 5 right) rhythm processing); red = attended processing / blue = unattended processing; solid line = deviant stimulus, dotted line = control stimulus. ERPs are collapsed over experimental groups. Scalp distributions relate to the deviant peaks.

**Discussion**

In the following section, we will discuss the behavioural results as well as the early negativity we observed for meter and for rhythm deviants, focussing on the effect the processing mode had on this component. Our discussion will close by reflecting on long-term training effect in auditory temporal processing.

*Behavioural data*

We found a difference in the discrimination rate between trained and untrained subjects. This result speaks for an effect of expertise in temporal processing due to long-term musical training and is consistent with previously reported behavioural data showing the superiority of musicians over non-musicians in the performance of temporal tasks such as the detection of timing variations (Jones, Jagacinski, Yee, Floyd, & Klapp, 1995), or the detection of tempo changes (Drake & Botte, 1993; Ehrle & Samson, 2005; Yee, Holleran, & Jones, 1994).

However, the percentage of correct answers clearly reveals a difference between the perception of meter and rhythm deviants, which is evidenced by a differential effect of expertise on the detection of the two temporal characteristics. The expertise effect was more pronounced on the meter condition than on the rhythm condition. Since this group difference was observed for the rhythm condition as well, although to a lesser extent, a ceiling effect on rhythm perception can be excluded. In other words, in untrained subjects rhythm perception performance is better than meter perception performance, while experts perform significantly better on both and bring meter perception to an equivalent level of performance as rhythm perception. We interpret this data in the sense that the ability to perceive meter is specifically necessary in music performance, whereas the ability to process rhythm seems to be also most relevant for untrained subjects, presumably in other perceptive domains of the auditory modality such as speech perception (van Zuijen, Sussman, Winkler, Naatanen, & Tervaniemi, 2005; Ramus, Hauser, Miller, Morris, & Mehler, 2000). There is evidence indicating that the perception of rhythmic aspects of speech highly facilitates speech comprehension (Friedrich, Kotz, Friederici, & Gunter, 2004). In addition, to the frequent exposure to speech rhythm, there is evidence in the literature

for rhythm perception as an "ancestral skill" of the auditory system with which to group incoming events. For example, primate research shows that tamarins are able to discriminate languages of different rhythm classes (Tincoff et al., 2005). While our data cannot disentangle the question of nature or nurture on temporal perception, it speaks for a clear proficiency of musically untrained subjects in temporal perception. Our data may further implicate a categorical difference between rhythm and meter in that superior perception of one element does not automatically allow for a good perception of the other.

*Electrophysiological data*

The electrophysiological analysis revealed a negative fronto-central deflection peaking at 150 ms time-locked to the onset of the deviant tone. The main purpose of this study was to identify differences between the perception of rhythm and meter deviants by manipulating the factors training and processing mode. After a general discussion of the event-related potential observed, we address the differences between meter and rhythm with respect to processing mode. Then we discuss the fact that we did not find any group differences in the negative deflection.

Negative deflections in the time window between 100 and 200 ms have been reported formerly in auditory experiments e.g. experiments related to pitch deviants within a melody (Nittono, Bito, Hayashi, Sakata, & Hori, 2000), to timbre discrimination (Meyer, Baumann, & Jancke, 2006), or changes of pitch and intensity (Snyder & Hillyard, 1976). Likewise, a negativity has been observed in the perception of auditory temporal deviants reported as an N1/P2 complex with a negativity peaking at 130 ms (Ford & Hillyard, 1981). An N2b component was reported in auditory omission conditions with a peak amplitude at 145-195 ms elicited by an auditory temporal omission condition (Raij, McEvoy, Makela, & Hari, 1997). These authors reported strong sensitivity of this component to attention, showing enhanced amplitudes in an attended processing condition as compared to an unattended processing condition. Similarly, another recent study reported an N150 related to temporal omission in an attended processing condition (Jongsma et al., 2005). The authors relate this component to the earlier reported N2b component and

interpret this effect as an orienting reaction due to temporal omission. Thus, in all of the above mentioned studies the negativity peaking between 100 and 200 ms time-locked to the stimulus onset was related to the perception of a temporal deviant. It was elicited in attended and unattended processing conditions and sometimes differentially manipulated by attention (Raij et al., 1997). Given the latency window of our component we are inclined to relate this finding to the N2b component associated with the perception of temporal deviants (Jongsma et al., 2005). However, this interpretation should be considered with caution since the scalp distributions of these components are slightly different to our findings. Our component displayed fronto-central negativity while the component reported by Jongsma and colleagues was centro-parietally localized.

One might alternatively interpret the N150 we observed as a mismatch negativity (MMN) - like component. It is well known that the mismatch negativity in the auditory modality is evoked in a specific pre-attentive oddball paradigm involving participants in the performance of a task such as reading a book or watching a silent movie (Naatanen, Pakarinen, Rinne, & Takegata, 2004; Ritter, De Sanctis, Molholm, Javitt, & Foxe, 2006; Gottselig, Brandeis, Hofer-Tinguely, Borbely, & Achermann, 2004; Restuccia, Della Marca G., Marra, Rubino, & Valeriani, 2005). However, an MMN-like component was recently reported when subjects focussed on another acoustic aspect of the stimulus material (Rinne, Sarkka, Degerman, Schroger, & Alho, 2006). In this experiment subjects were listening to pitch differences while the MMN was elicited by intensity deviants. The observed component was described as MMN-like on the basis of its latency window as well as due to its characteristic scalp distribution displaying a fronto-central negativity. This distribution is clearly distinct from the central negativity commonly observed for the N1 component. Similarly, Sabri et al. have reported an MMN in the time range of 110 and 170 ms elicited by pitch deviants while subjects performed an auditory duration task (Sabri, Liebenthal, Waldron, Medler, & Binder, 2006).

The paradigm adopted in our experiment was comparable with the paradigm used in those studies. Furthermore, the component elicited in the unattended processing condition showed a scalp distribution similar to that of the MMN (Shalgi & Deouell, 2007; Ritter et al., 2006)[1]. Thus, the latency of the negativity as well as the scalp

distribution observed in our experiment might speak for an MMN-like component. However, since we did not use the standard MMN-paradigm we choose to refere to our component as a N150. The N150 was found for the temporal deviants compared to the standard tones. This finding indicates that the perception of temporal changes induces a more extensive neuronal activity and, furthermore, that this effect is possibly due to an orienting reaction. In the following section we discuss the observed between meter and rhythm deviants differences in this component.

*Differences between meter and rhythm processing*

Most evident is the difference in negativity observed between rhythm and meter deviants. The rhythm deviants elicited an N150 with higher amplitudes compared to the N150 elicited by the meter deviant. However, this difference in amplitude cannot be attributed to the temporal characteristic of our experimental stimuli solely, but could be influenced by the difference in intensity of the tones. A meter change was always time-locked to an accented tone, while a rhythm change was unaccented. The most relevant difference between meter and rhythm-related ERPs is to be found within the differential effects that the processing mode had on the observed component.

The rhythm deviants elicited an N150 independent of processing mode. However, the meter deviant only elicited a negative deflection in the attended processing condition. One might argue that this effect is due to a greater salience of rhythm as compared with meter changes. However, musicians performed just as well in the meter condition as compared to the rhythm condition in both groups. Since this group difference did not show in the electrophysiological data, we are of the opinion that this attention related difference could speaks for the existence of different neurofunctional mechanisms underlying the processing of meter and rhythm. More precisely, the meter perception could differ from the rhythm percept in the time course of its perception. Early theoretical models have already suggested that meter perception relies on Gestalt phenomena and hence requires the integration of several sequential auditory cues and subsequent reinterpretation (Cooper et al., 1960). Therefore, it is plausible that the perception of meter in the unattended processing

condition takes place later than 150 ms after the stimulus onset. Contrarily, in the attended processing condition, when listerners focussed on temporal changes, we observed a clear N150 for meter changes. One must therefore infer that this attention dependent negativity in the context of the present study reflects a percept of meter. This poses a question on the suggested difference between meter and rhythm processing.

Vuust and colleagues (2005) found an effect of meter perception reflected by a negativity between 100 and 150 ms after stimulus onset in an unattended processing condition. This negativity was elicited by two different temporal deviants and displayed increased amplitudes for the stronger, as compared to the weaker, deviant. We are of the opinion that our finding is only at first sight contradictory to the finding reported by Vuust and colleagues. Rather, this difference is due to the specific construction of the stimuli in that study. The weaker manipulation in the study by Vuust et al. was a syncope, which can be described as a rhythm change without affecting the underlying metric pattern. It is specific for the meter percept that it is not affected by short irregularity, such as a syncope (Meyer, 1956). We suggest that the component observed by Vuust et al. could be comparable to the negativity we observed in the rhythm condition. With reference to the reported strong violation, Vuust and colleagues mention the introduction of a beat which is incongruent with the underlying metrical grid, making it seem as if the music were to "stumble" (Vuust et al. 2005, p. 561). This violation, however, strongly differs from our metric manipulation. Vuust et al. delayed an expected beat, a manipulation which could also be considered a caesura. This is a musical terminus for a small delay of an expected tone used as an effect in musical interpretation. Thus, one would expect that this effect does not affect the percept of the underlying meter. In our study, we took a different approach by removing a tone. The deviant tone appeared exactly at the time expected, while the dynamic characteristic of this tone was unexpected (Toivainen et al., 2004). Thus, the two metrical manipulations are crucially different in their nature, which most likely causes the observed electrophysiological difference. In sum, the fact that the metrical manipulation in our data failed to elicit an N150 in the unattended processing condition even in high performing trained subjects, while the rhythm manipulation elicited an N150, might sustain the notion of a difference

between meter and rhythm processing. However, this effect of processing mode on the processing of the meter needs to be further investigated.

*No differences between trained and untrained subjects*

Interestingly, we did not find differences in the N150 component between musicians and non-musicians. This electrophysiological finding suggests that musicians and nonmusicians perceive meter and rhythm changes in the same way.

Several authors investigated training effects on auditory processing. There is evidence of short-term training effects in components of 100-200 ms latency which were elicited by melodic changes, different voice-onset times, or frequency changes (Bosnyak, Eaton, & Roberts, 2004; Tervaniemi, Rytkonen, Schroger, Ilmoniemi, & Naatanen, 2001; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). Long-term training effects have been found by comparing musicians and nonmusicians in the perception of pure tones and pitch grouping (Shahin, Bosnyak, Trainor, & Roberts, 2003; van Zuijen, Sussman, Winkler, Naatanen, & Tervaniemi, 2004). More specifically, a long-term training effect on temporal processing has been found by Jongsma and colleagues reporting higher N150 amplitudes time-locked to an auditory temporal omission (Jongsma et al., 2005).

Yet, there are studies that do not find training-effects in auditory processing of pure tones, familiar or unfamiliar chords, or the violation of temporal irregularity (Neuloh & Curio, 2004; van Zuijen et al., 2005; Lutkenhoner, Seither-Preisler, & Seither, 2006). Thus, it appears that the literature on this matter is yet inconclusive and that possible training differences might occur, although not in every aspect of temporal processing. In accordance with van Zuijen and colleagues, we argue that the unobserved group differences on the N150 in either the unattended or attended temporal processing clearly speaks for a functional relevance of these perceptive processes, not only pertinent to musically trained subjects. This relevance is possibly related to the perception of speech that highly relies on temporal segmentation processes.

Contrary to this electrophysiological finding, behavioural results of our study clearly show differences in the performance rate between the two groups. Musicians performed significantly better on the detection task, particularly in the detection of

meter deviants. These findings seem conflicting. That musicians are more proficient in meter perception as compared with non-musicians has been documented earlier (Ehrle et al., 2005). In our opinion, this discrepancy between our electrophysiological and our behavioural data is yet another indication that the electrophysiological correlate of the meter percept is not reflected in the early latency window we analysed. This hypothesis is in line with a considerable amount of research reporting short-term training effects reflected in late ERP, for example, an enhanced P2-amplitude or P2m dipole moment (Atienza, Cantero, & Dominguez-Marin, 2002; Kuriki, Kanda, & Hirata, 2006; Reinke, He, Wang, & Alain, 2003; Shahin, Roberts, Pantev, Trainor, & Ross, 2005). On the basis of these findings we speculate that the behavioural differences observed in the present experiment might be reflected in electrophysiological processes within a later latency window. However, this issue needs further investigation.

One difference we observed between musicians and non-musicians lies in the scalp potential distributions related to the attended processing of meter deviants. While musicians showed centro-frontal negativity, the negativity was more lateralized to right frontal regions in non-musicians[2]. Although this lateralization difference is not sustained statistically in a source estimation analysis, we may speculate that this differential effect relates to a well described difference between trained and untrained subjects. There is increasing evidence that one of the main characteristics of the musically trained brain lies within a more distributed network underlying the processing of acoustic stimuli in musicians, specifically the interaction between the hemispheres. Anatomical as well as functional evidence supports this concept (Gaser & Schlaug, 2003; Jancke, 2002; Schlaug, Jancke, Huang, Staiger, & Steinmetz, 1995; Schlaug, Jancke, Huang, & Steinmetz, 1995).

**Summary**

Our study reveals an electrophysiological correlate of auditory rhythm perception in the form of a negative deflection in the latency window of 100-200 ms. This deflection is present in both attended and unattended processing. Meter deviants only elicit a similar negative deflection in the attended processing condition of our study.

Our data provide evidence on one hand of a categorical difference between rhythm and meter, as evidenced by this effect of processing mode as well as by a differential effect of expertise on the perception of meter and rhythm. On the other hand, this data might be interpreted as evidence for rhythm perception as an "ancestral skill" of the auditory system with which to group incoming events.

Acknowledgments

During the preparation of this manuscript E.G. was supported by the Swiss National Sciences Foundation (SNF) 3200B0-105877 and the "Fonds zur Förderung des Akademischen Nachwuchses". With thank Elsa Sägesser and Marcus Cheetham for very helpful comments on former versions of this manuscript.

Reference List

Annett, M. (1992). Five tests of hand skill. *Cortex, 28,* 583-600.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The Time Course of Neural Changes Underlying Auditory Perceptual Learning. *Learning and Memory, 9,* 138-150.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex, 14,* 1088-1099.

Chatrian, G. E., lettich, E., & Nelson, P. L. (1988). Modified nomenclature for the "10%" electrode system. *Journal of Clinical Neurophysiology, 5,* 183-186.

Cooper, G. W. & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.

Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia, 42,* 868-877.

Drake, C. & Botte, M. C. (1993). Tempo sensitivity in auditory sequences - evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277-286.

Drake, C. & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343-378.

Ehrle, N. & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition, 58,* 133-147.

Essens, P. J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics, 40,* 69-73.

Ford, J. M. & Hillyard, S. A. (1981). Event-Related Potentials (Erps) to Interruptions of A Steady Rhythm. *Psychophysiology, 18,* 322-330.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience, 16,* 541-552.

Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *the Journal of Neuroscience, 23,* 9240-9245.

Gordon, E. E. Advanced Measures of Music Audition. [Audiovisual Material] Chicago: Publications Inc. , 1984.

Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbely, A. A., & Achermann, P. (2004). Human central auditory plasticity associated with tone sequence learning. *Learning & Memory, 11,* 162-171.

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology-Human Perception and Performance, 30,* 956-974.

Ibbotson, N. r. & Morton, J. (1981). Rhythm and dominance. *Cognition, 9,* 135.

Jancke, L. (2002). What is special about the brains of musicians? *Neuroreport, 13,* 741-742.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of Attentional Flexibility in Listening to Polyrhythmic Patterns. *Journal of Experimental Psychology-Human Perception and Performance, 21,* 293-307.

Jongsma, M. L., Eichele, T., Quian, Q. R., Jenks, K. M., Desain, P., Honing, H. et al. (2005). Expectancy effects on omission evoked potentials in musicians and non-musicians. *Psychophysiology, 42,* 191-201.

Kuck, H., Grossbach, M., Bangert, M., & Altenmuller, E. (2003). Brain processing of meter and rhythm in music. Electrophysiological evidence of a common network. *Annals of the New York Academy of Sciences, 999,* 244-253.

Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *Journal of Neuroscience, 26,* 4046-4053.

Lehrdahl, F. & Jackendoff, R. (1981). On lthe theory of grouping and meter. *The musical Quarterly, 67,* 479-506.

Lerdahl, F. & Jackendoff, R. (1983). An Overview of Hierarchical Structure in Music. *Music Perception, 1,* 229-252.

Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain, 121 ( Pt 10),* 1853-1867.

Lutkenhoner, B., Seither-Preisler, A., & Seither, S. (2006). Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians. *Neuroimage, 30,* 927-937.

MacMillan N.A. (1991). *Detection theory: A users' guide*. (1 ed.) (vols. 1) Cambridge: Cambridge University Press.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Meyer, M., Baumann, S., & Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage., 32,* 1510-1523.

Midorikawa, A., Kawamura, M., & Kezuka, M. (2003). Musical alexia for rhythm notation: A discrepancy between pitch and rhythm. *Neurocase, 9,* 232-238.

Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition, 56,* 36-42.

Naatanen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology, 115,* 140-144.

Neuloh, G. & Curio, G. (2004). Does familiarity facilitate the cortical processing of music sounds? *Neuroreport, 15,* 2471-2475.

Nittono, H., Bito, T., Hayashi, M., Sakata, S., & Hori, T. (2000). Event-related potentials elicited by wrong terminal notes: effects of temporal disruption. *Biological Psychology, 52,* 1-16.

Penhune, V. B., Zatorre, R. J., & Feindel, W. H. (1999). The role of auditory cortex in retention of rhythmic patterns as studied in patients with temporal lobe removals including Heschl's gyrus. *Neuropsychologia, 37,* 315-331.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100,* 1-32.

Peretz, I. & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: a neuropsychological perspective. *Quarterly Journal of Experimental Psychology, 46,* 301-325.

Peretz, I. & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56,* 89-114.

Phillips-Silver, J. & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition, 105,* 533-546.

Piccirilli, M., Sciarma, T., & Luzzi, S. (2000). Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery, and Psychiatry, 69,* 541-545.

Povel, D. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411-440.

Raij, T., McEvoy, L., Makela, J. P., & Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Research, 745,* 134-143.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288,* 349-351.

Reinke, K. S., He, Y., Wang, C. H., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research, 17,* 781-791.

Restuccia, D., Della Marca G., Marra, C., Rubino, M., & Valeriani, M. (2005). Attentional load of the primary task influences the frontal but not the temporal generators of mismatch negativity. *Cognitive Brain Research, 25,* 891-899.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. *Neuroimage., 16,* 169-176.

Rinne, T., Sarkka, A., Degerman, A., Schroger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research, 1077,* 135-143.

Ritter, W., De Sanctis, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology, 43,* 423-430.

Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A., & Binder, J. R. (2006). Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. *Journal of Cognitive Neuroscience, 18,* 689-700.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K. et al. (1999). Neural representation of a rhythm depends on its interval ratio. *the Journal of Neuroscience, 19,* 10074-10081.

Schachter, C. (1986). Aspects of meter. *Music Forum, 6,* 1-59.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia, 33,* 1047-1055.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267,* 699-701.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience, 23,* 5545-5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport, 16,* 1781-1785.

Shalgi, S. & Deouell, L. Y. (2007). Direct evidence for differential roles of temporal and frontal components of auditory change detection. *Neuropsychologia*.

Snyder, E. & Hillyard, S. A. (1976). Long-Latency Evoked-Potentials to Irrelevant, Deviant Stimuli. *Behavioral Biology, 16,* 319-331.

Tervaniemi, M., Rytkonen, M., Schroger, E., Ilmoniemi, R. J., & Naatanen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory, 8,* 295-300.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. *Developmental Science, 8,* 26-35.

Toivainen, P. & Eerola, T. (2004). The role of accent periodicies in meter induction: a classification study. *Proceedings of the 8th International Conference on Music Perception & Cognition,* 422-425.

Trehub, S. E. & Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition, 100,* 73-99.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing, 22,* 79-90.

van Zuijen, T. L., Sussman, E., Winkler, I., Naatanen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds--an event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience, 16,* 331-338.

van Zuijen, T. L., Sussman, E., Winkler, I., Naatanen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity-a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research, 23,* 270-276.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A. et al. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage, 24,* 560-564.

Wilson, S. J., Pressing, J. L., & Wales, R. J. (2002). Modelling rhythmic function in a musician post-stroke. *Neuropsychologia, 40,* 1494-1505.

Yee, W., Holleran, S., & Jones, M. R. (1994). Sensitivity to Event Timing in Regular and Irregular Sequences - Influences of Musical Skill. *Perception & Psychophysics, 56,* 461-471.

# VI General Discussion

The findings of each individual study have been discussed in the specific discussion section of the manuscripts. The following chapter will briefly summarize the findings and interpretation. Subsequently, the most important findings will be integrated in the context of the general research questions stated at the beginning of the present thesis.

The *first study* investigated the neural correlates of rhythm processing in speech perception. The data revealed that explicit perception of exaggerated German speech rhythm relies on posterior superior temporal brain structures (pSTG) of the right hemisphere, while implicit perception of the same stimuli relies on pSTG of the left hemisphere. These data are one of the first findings to suggest a differential lateralization of auditory cortex activation depending on the processing mode. The nature of this observed top-down processing mechanisms must be investigated further using other stimulus material as well as other processing modes.
The specific network involved in the performance of a speech rhythm task shed a new light on the function of the supplementary motor area and sustained hypothesis about an insular function in auditory timing perception. Until now, the supplementary motor area had been specifically associated with motor processes. Our results lead to the assumption of a more general involvement of the SMA in timing, and furthermore, that this function is equally exerted in productive as well as in perceptive processes. This finding is in congruence with timing studies in the visual domain (Coull, 2004; Macar, Anton, Bonnet, & Vidal, 2004; Livesey, Wall, & Smith, 2007). The exact temporal function of the SMA in timing perception remains to be investigated.

The *second study* investigated the neural correlates of speech melody processing. The results were able to replicate former findings on the involvement of the planum temporale as well as the inferior frontal gyrus in the perception of speech melody. Furthermore, the present results were able to show that the brain area critically involved in the interaction between semantic and prosodic speech perception is the posterior part of the middle temporal gyrus as well as the inferior frontal gyrus. With

respect to a differential effect of processing mode (attentional vs. unattentional), the results indicated an interaction between the nature of the prosodic manipulation and the processing mode in secondary auditory cortex areas, namely the planum temporale and the posterior middle temporal gyrus. Depending on the processing condition, greater activity was either observed for the congruency condition of a specific stimulus characteristic or for the non-congruency condition. This has implications for design construction in future experiments and should be subject to further investigations.

The *third study* investigated the electrophysiological correlates of meter and rhythm perception in an explicit and an implicit processing condition. The data revealed an early negative deflection in the explicit processing condition for temporal changes of both categories compared to standard tones. In the implicit processing condition only rhythm changes elicited this negative deflection. These findings suggest that meter and rhythm perception might differ with respect to the time course of perception, thus suggesting that both processing mechanisms rely on different neurophysiological mechanisms.

*Neural correlates of suprasegmental auditory processing*: The goal of the present work was to investigate the neural correlates of suprasegmental auditory processing by investigating rhythm and melody perception. With respect to rhythm perception the question of hemispheric lateralization was of particular interest. Results indicated an involvement of secondary auditory cortex areas, namely the posterior superior temporal gyrus, in the perception of speech melody as well as in the perception of exaggerated German speech rhythm. The posterior superior temporal gyrus has previously been associated with suprasegmental auditory processing, specifically with speech melody perception (Meyer et al., 2002). However, the perception of acoustic rhythm in this brain area has not been observed before. The specific function of the pSTG has been described as encompassing acoustic spectrotemporal processing (Griffiths et al., 2002) making this brain region a core region in auditory functional lateralization (Zatorre et al., 2001; Poeppel, 2003). The present findings indicate that the specific function of the pSTG in suprasegmental auditory processing is not restricted to spectral processing, nor processing of rapidly changing temporal

cues but rather includes the analysis of temporal auditory signals on a suprasegmental level, as well.

As mentioned in the introduction, the neural correlates of prosody processing in speech have been investigated extensively in the past sustaining the assumption that the right hemisphere in particular is involved in suprasegmental, or speech melody, processing (Meyer et al., 2002; Belin, Zatorre, & Ahad, 2002). At least in rhythm perception, the present data however, indicate an *effect of processing mode on the functional lateralization of suprasegmental auditory processing*. The specific experimental paradigms used in all three experiments allowed for comparisons between explicit (attentional) and implicit (unattentional) processing of the stimulus characteristic of interest. The speech rhythm experiment revealed a lateralization difference in that explicit processing of speech rhythm showed activity in the right pSTG, while implicit processing of speech rhythm was associated with activity in the left pSTG. Comparable top-down modulation effects on sensory cortical areas have earlier been suggested (Noesselt et al., 2003; Jancke, Mirzazade, & Shah, 1999a). However, the present findings even speak for a modulation effect on functional lateralization suggesting a dynamic interplay between corresponding cortical areas of both hemispheres (Scheich et al., 2007; Tervaniemi et al., 2003). Therefore, the present data suggest that it is necessary to consider the processing condition in addition to the specific stimulus characteristic in order to truly describe neural correlates of auditory processing.

**Conclusion**

The present work has extended the knowledge regarding the neural correlates of suprasegmental auditory processing, namely the posterior part of the superior temporal gyrus (planum temporale), as a brain area critically involved in temporal auditory processing. Moreover, top-down modulatory effects on the secondary auditory cortex have been found.

Reference List

Aasland, W. A. & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: a comparison of processing by left- and right-hemisphere brain-damaged individuals. Brain Lang, 87, 385-399.

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. Neuroimage, 35, 1192-1200.

Ashburner, J. & Fristson, K. J. (1999). Nonlinear Spatial Normalization using Basis Functions. In (pp. 254-266).

Astesano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. Brain Res.Cogn Brain Res., 18, 172-184.

Barrett, D. J. & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. Neuroimage, 32, 968-977.

Baum, S. R. & Dwivedi, V. D. (2003). Sensitivity to prosodic structure in left- and right-hemisphere-damaged individuals. Brain Lang, 87, 278-289.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. Brain Res.Cogn Brain Res., 13, 17-26.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. Neuroimage., 10, 417-429.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat.Neurosci., 7, 295-301.

Brechmann, A. & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cerebral Cortex, 15, 578-587.

Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde. Leipzig: Barth.

Brown, M. Ch. (2003). Audition. In L.R.Squire, F. E. Bloom, S. K. McConnell, J. L. Roberts, N. C. Spitzer, & Zigmond M.J. (Eds.), Fundamental Neuroscience (2. ed., pp. 699-726). San Diego: Academic Press.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. Brain Res.Cogn Brain Res., 20, 363-375.

Bussmann, H. (1990). Lexikon der Sprachwissenschaft. (2. ed.) Stuttgart: Kröner.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson.Med., 39, 855-864.

Chen, C. K., Chiueh, T. D., & Chen, J. H. (1999). Active cancellation system of acoustic noise in MR imaging. IEEE Trans.Biomed.Eng, 46, 186-191.

Coull, J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. Brain Res.Cogn Brain Res., 21, 216-226.

Cremillieux, Y., Wheeler-Kingshott, C. A., Briguet, A., & Doran, S. J. (1997). STEAM-Burst: a single-shot, multi-slice imaging sequence without rapid gradient switching. Magn Reson.Med., 38, 645-652.

Crystal, D. (1975). The English tone of voice. Essays in intonation, prosody and paralanguage. London. London.

Cutler, A. (1994). Segmentation Problems, Rhythmic Solutions. Lingua, 92, 81-104.

Darwin, C. (1871). The descent of man and selection in realtion to sex. London: John Murray.

Davis, M. H. & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. J.Neurosci., 23, 3423-3431.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. Journal of Neurophysiology, 2, 494-499.

Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. Neuropsychologia, 42, 868-877.

Eckert, M. A., Leonard, C. M., Molloy, E. A., Blumenthal, J., Zijdenbos, A., & Giedd, J. N. (2002). The epigenesis of planum temporale asymmetry in twins. Cereb.Cortex, 12, 749-755.

Eckstein, K. & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. Brain Res.Cogn Brain Res., 25, 130-143.

Eckstein, K. & Friederici, A. D. (2006). It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. J.Cogn Neurosci., 18, 1696-1711.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K. et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage., 25, 1325-1335.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. Cognition, 100, 173-215.

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., & Bowtell, R. W. (2000). Sound-level measurements and calculations of safe noise dosage during EPI at 3 T. J.Magn Reson.Imaging, 12, 157-163.

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C. et al. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology, 63, 1640-1646.

Fox, P. T. & Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc.Natl.Acad.Sci.U.S.A, 83, 1140-1144.

Frahm, J., Fransson, P., & Krüger, G. (1999). Magnetic resonance imaging of human brain function. In U.Windhorst & H. Johansson (Eds.), Modern echniques in neuroscience research ( Heiderlberg: Springer Verlag.

Friederici, A. D. & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang, 89, 267-276.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. Brain Res.Cogn Brain Res., 20, 300-308.

Friston, K. J. (2004). Experimental design and statistical parametric mapping. In R.S.Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. Ashburner, & W. Penny (Eds.), 632 (pp. -599). Oxford: Academic Press.

Galaburda, A. & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. J.Comp Neurol., 190, 597-610.

Grice, M., Baumann, St., & Benzmüller, R. (2005). German INtonation in Autosegmental-Metrical PHonology. In S.-A.Jun (Ed.), The Phonology of Intonation and PHrasing ( Oxford: University Press.

Griffiths, T. D., Kumar, S., Warren, J. D., Stewart, L., Stephan, K. E., & Friston, K. J. (2007). Approaches to the cortical analysis of auditory objects. Hear.Res., 229, 46-53.

Griffiths, T. D. & Warren, J. D. (2002). The planum temporale as a computational hub. Trends Neurosci., 25, 348-353.

Guiraud, J., Besle, J., Arnold, L., Boyle, P., Giard, M. H., Bertrand, O. et al. (2007). Evidence of a tonotopic organization of the auditory cortex in cochlear implant users. J.Neurosci., 27, 7838-7846.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. J.Cogn Neurosci., 16, 1302-1318.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R. et al. (1999). "Sparse" temporal sampling in auditory fMRI. Hum.Brain Mapp., 7, 213-223.

Hattori, Y., Fukatsu, H., & Ishigaki, T. (2007). Measurement and evaluation of the acoustic noise of a 3 Tesla MR scanner. Nagoya J.Med.Sci., 69, 23-28.

Heiervang, E., Hugdahl, K., Steinmetz, H., Inge, S. A., Stevenson, J., Lund, A. et al. (2000). Planum temporale, planum parietale and dichotic listening in dyslexia. Neuropsychologia, 38, 1704-1713.

Herdener, M., Esposito, F., Di Salle, F., Lehmann, C., Bach, D. R., Scheffler, K. et al. (2007). BOLD correlates of edge detection in human auditory cortex. Neuroimage, 36, 194-201.

Hesling, I., Clement, S., Bordessoules, M., & Allard, M. (2005). Cerebral mechanisms of prosodic integration: evidence from connected speech. Neuroimage., 24, 937-947.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. Science, 182, 177-180.

Hirst, D. J. (2005). Form and function in the representation of speech prosody. Speech Communication, 46, 334-347.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. Neurosci.Res., 54, 276-280.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M. et al. (2000). Auditory cortex on the human posterior superior temporal gyrus. J.Comp Neurol., 416, 79-92.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang, 76, 227-252.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Hum.Brain Mapp., 26, 128-138.

Inui, K., Okamoto, H., Miki, K., Gunji, A., & Kakigi, R. (2006). Serial and Parallel Processing in the Human Auditory Cortex: A Magnetoencephalographic Study. Cereb.Cortex, 16, 18-30.

Jancke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neurosci.Lett., 266, 125-128.

Jancke, L. & Steinmetz, H. (1993). Auditory lateralization and planum temporale asymmetry. Neuroreport, 5, 169-172.

Jaspers, H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalogr.Clin.Neurophysiol., 10, 371-375.

Josephs, O., Turner, R., & Friston, K. (1997). Even-related fMRI. Human Brain Mapping, 5, 248.

Josse, G., Mazoyer, B., Crivello, F., & Tzourio-Mazoyer, N. (2003). Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. Brain Res.Cogn Brain Res., 18, 1-14.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iraguei, V. et al. (2000). Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 37, 163-178.

Jusczyk, P. W. (1999). How infants begin to extract words From speech. Trends in Cognitive Sciences, 3, 323-328.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. Neuroimage., 14, 1402-1408.

Kotz, S. A. & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. Brain Res., 1151, 107-118.

Krumbholz, K., Schonwiesner, M., Rubsamen, R., Zilles, K., Fink, G. R., & von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. European Journal of Neuroscience, 21, 230-238.

Langers, D. R., Backes, W. H., & van Dijk, P. (2007). Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. Neuroimage, 34, 264-273.

Leonard, C. M., Puranik, C., Kuldau, J. M., & Lombardino, L. J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? Cereb.Cortex, 8, 397-406.

Lerdahl, F. (2001). The Sounds of Poetry Viewed as Music. Annals of the New York Academy of Sciences, 930, 337-354.

Livesey, A. C., Wall, M. B., & Smith, A. T. (2007). Time perception: manipulation of task difficulty dissociates clock functions from other cognitive demands. Neuropsychologia, 45, 321-331.

Luck, St. J. (2005). Ten simple rules of designing ERP experiments. In T.C.Handy (Ed.), Even.related potentials: A methods handbook (pp. 17-32). Cambridge: MIT Press.

Luders, E., Gaser, C., Jancke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. Neuroimage, 22, 656-664.

Macar, F., Anton, J. L., Bonnet, M., & Vidal, F. (2004). Timing functions of the supplementary motor area: an event-related fMRI study. Brain Res.Cogn Brain Res., 21, 206-215.

Mansfield, P., Glover, P. M., & Beaumont, J. (1998). Sound generation in gradient coil structures for MRI. Magn Reson.Med., 39, 539-550.

McJury, M. & Shellock, F. G. (2000). Auditory noise associated with MR procedures: a review. J.Magn Reson.Imaging, 12, 37-45.

Merker, B. (2000). The origins of music. In N.L.Wallin, B. Merker, & S. Brown (Eds.), Synchronous chorusing and human origins (pp. 315-327). Cambridge, Mass: The MIT Press.

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp., 17, 73-88.

Moelker, A. & Pattynama, P. M. (2003). Acoustic noise concerns in functional magnetic resonance imaging. Hum.Brain Mapp., 20, 123-141.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage., 13, 684-701.

Morosan, P., Schleicher, A., Amunts, K., & Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. Anat.Embryol.(Berl), 210, 401-406.

Murakami, S. & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J.Physiol, 575, 925-936.

Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication, 41, 233-243.

Nishimura, M., Shirasawa, H., Kaizo, H., & Song, W. J. (2007). New field with tonotopic organization in guinea pig auditory cortex. J.Neurophysiol., 97, 927-932.

Noesselt, T., Shah, N. J., & Jancke, L. (2003). Top-down and bottom-up modulation of language related areas - An fMRI Study. BMC.Neurosci., 4, 13.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc.Natl.Acad.Sci.U.S.A, 87, 9868-9872.

Ozaki, I. & Hashimoto, I. (2007). Human tonotopic maps and their rapid task-related changes studied by magnetic source imaging. Can.J.Neurol.Sci., 34, 146-153.

Patel, A. D. (2003). Rhythm in language and music: parallels and differences. Ann.N.Y.Acad.Sci., 999, 140-143.

Patel, A. D. & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. Cognition, 87, B35-B45.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang, 61, 123-144.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. Neuron, 36, 767-776.

Paul, L. K., Lancker-Sidtis, D., Schieffer, B., Dietrich, R., & Brown, W. S. (2003). Communicative deficits in agenesis of the corpus callosum: nonliteral language and affective prosody. Brain Lang, 85, 313-324.

Pell, M. D. (2005). Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang.

Penhune, V. B., Zatorre, R. J., & Feindel, W. H. (1999). The role of auditory cortex in retention of rhythmic patterns as studied in patients with temporal lobe removals including Heschl's gyrus. Neuropsychologia, 37, 315-331.

Peretz, I. & Zatorre, R. J. (2005). Brain organization for music processing. Annu.Rev.Psychol., 56, 89-114.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41, 245-255.

Poline, J. B. (2003). Contrasts and classical inference. In R.S.J.Frackowiak, K. J. Friston, C. Frith, R. Dolan, K. J. Friston, C. J. Price, S. Zeki, J. Ashburner, & W. D. Penny (Eds.), Human Brain Function (2nd ed., Academic Press.

Preis, S., Jancke, L., Schmitz-Hillebrecht, J., & Steinmetz, H. (1999). Child age and planum temporale asymmetry. Brain Cogn, 40, 441-452.

Purves, D., Ditzpatrick, D., Williams, S. M., McNamara, J. O., Augustine, G. J., Katz L.C. et al. (2001). Neuroscience. (2nd edition ed.) Sunderland: Sinauer Associates.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J. et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage, 13, 669-683.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. Neuroimage., 16, 169-176.

Rinne, T., Kirjavainen, S., Salonen, O., Degerman, A., Kang, X., Woods, D. L. et al. (2007). Distributed cortical networks for focused auditory attention and distraction. Neurosci.Lett., 416, 247-251.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat.Neurosci., 2, 1131-1136.

Scheich, H., Baumgart, F., Gaschler-Markefski, B., Tegeler, C., Tempelmann, C., Heinze, H. J. et al. (1998). Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. Eur.J.Neurosci., 10, 803-809.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & OHl, F. W. (2007). The Cognitive Auditory Cortex: Task-Specificity of Stmulus REpresentations. Hearing Research.

Schmidt, C. F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., & Jancke, L. (2007). Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. Human Brain Mapping, doi:10.1002/hbm.20372.

Schonwiesner, M., Krumbholz, K., Rubsamen, R., Fink, G. R., & von Cramon, D. Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. Cereb.Cortex, 17, 492-499.

Shah, N. J., Jancke, L., Grosse-Ruyken, M. L., & Muller-Gartner, H. W. (1999). Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J.Magn Reson.Imaging, 9, 19-25.

Shapleske, J., Rossell, S. L., Woodruff, P. W., & David, A. S. (1999). The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. Brain Res.Brain Res.Rev., 29, 26-49.

Shellock, F. G., Ziarati, M., Atkinson, D., & Chen, D. Y. (1998). Determination of gradient magnetic field-induced acoustic noise associated with the use of echo planar and three-dimensional, fast spin echo techniques. J.Magn Reson.Imaging, 8, 1154-1157.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, c., Price, P. et al. (1992). ToBI: a standard for labeling English prosody. [92((2)], 867-870. Banff, Canada. Proc. ICSLP.

Ref Type: Serial (Book,Monograph)

Steinmetz, H., Rademacher, J., Huang, Y. X., Hefter, H., Zilles, K., Thron, A. et al. (1989). Cerebral asymmetry: MR planimetry of the human planum temporale. J.Comput.Assist.Tomogr., 13, 996-1005.

Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y. X., Thron, A., & Zilles, K. (1990). Total surface of temporoparietal intrasylvian cortex: diverging left-right asymmetries. Brain Lang, 39, 357-372.

Takayanagi, M. & Ojima, H. (2006). Microtopography of the dual corticothalamic projections originating from domains along the frequency axis of the cat primary auditory cortex. Neuroscience, 142, 769-780.

Talairach, J. & Tournoux, P. (1988). Co-Planar Sterotaxic Atlas of the Human Brain. Stuttgart: Thieme.

Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev., 43, 231-246.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage., 15, 273-289.

Warren, J. D. & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J.Neurosci., 23, 5799-5804.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. Proc.Natl.Acad.Sci.U.S.A, 100, 10038-10042.

Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantifying variability in the planum temporale: a probability map. Cereb.Cortex, 9, 392-405.

Wilson, S. J., Pressing, J. L., & Wales, R. J. (2002). Modelling rhythmic function in a musician post-stroke. Neuropsychologia, 40, 1494-1505.

Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci., 20, 2447-2456.

Zatorre, R. J. & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. Cereb.Cortex, 11, 946-953.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? Nat.Neurosci., 5, 905-909.

Zatorre, R. J. & Penhune, V. B. (2001). Spatial localization after excision of human auditory cortex. J.Neurosci., 21, 6321-6328.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. Neuroimage.

Zilles, K., Palomero-Gallagher, N., Grefkes, C., Scheperjans, F., Boy, C., Amunts, K. et al. (2002). Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. Eur.Neuropsychopharmacol., 12, 587-599.

# Appendix

## List of Abbreviations

| | |
|---|---|
| AC | auditory cortex |
| ANOVA | analysis of variance |
| BOLD | blood oxygen dependence level |
| EEG | electroencephalography |
| EPI | echo planar imaging |
| ER | error rate |
| ERP | event-related potential |
| fMRI | functional magnetic resonance imaging |
| FOP | frontal operculum |
| FOP | insular cortex |
| GLM | general linear model |
| HG | Heschl's gyrus |
| HRF | hemodynamic response function |
| IFG | temporal gyrus |
| MMN | mismatch negativity |
| MNI | Montreal Neurological Institute |
| MR | magnetic resonance |
| MTG | middle Superior temporal gyrus |
| POP | parietal operculum |
| PP | planum polare |
| PR | performance rate |
| PT | planum temporale |
| ROP | rolandic operculum |
| RT | reaction time |
| SMA | supplementary motor gyrus |
| SMG | supramarginal gyrus |
| STG | inferior frontal gyrus |

## Curiculum vitae

### Personal

| | |
|---|---|
| Full name | Eveline Geiser |
| Data of birth | 04.09.1977 in Männedorf / ZH, Switzerland |

### Education

| | |
|---|---|
| 1984-1990 | Primary School in Hausen am Albis / ZH, Switzerland |
| 1990-1996 | Secondary School in Zurich Wiedikon, Switzerland |
| 1994/1995 | Highschool in Pennsylvania, USA |
| 1998 -2003 | Studies in psychology and business economics at the University of Zurich (Master of Science), Switzerland |
| 2003/2004 | Diploma thesis at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany |
| 2005 - 2007 | PhD-student at the University of Zurich, Department of Psychology, Division Neuropsychology, Switzerland |
| since Sept. 2008 | Post-doctoral fellow at the Department for Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA. USA |

### Peer-Reviewed Publications

Geiser, E., Zaehle, T.,, Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by meter processing: an fmri study. *Journal of Cognitive Neuroscience, 20(3),* 179-90.

Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2008). Electrophysiological correlates of meter and rhythm processing in music perception. *Cortex,* doi:10.1016/j.cortex.2007.09.010.

Zaehle, T., Geiser, E., Alter, K., Jancke, L. & Meyer, M. (2008). Segmental processing in the human dorsal auditory stream. *Brain Research,. 1220,* 179-90.

Schmidt, C.F., Meyer, M., Zaehle, T., Geiser, E., Boesiger, P., & Jancke, L. (2008). Silent and continuous scanning differentially modulate activation in an auditory language comprehension task. *Human Brain Mapping. 29(1),* 46-56.